UNIVERSITY OF GRONINGEN

JOHANN BERNOULLI INSTITUTE FOR MATHEMATICS AND COMPUTER SCIENCE

UNIVERSITY OF LEÓN

DEPARTMENT OF ELECTRICAL, SYSTEMS AND AUTOMATIC ENGINEERING

OBJECT RECOGNITION TECHNIQUES IN REAL APPLICATIONS

A dissertation supervised by

PROF. DR. ENRIQUE ALEGRE GUTIÉRREZ,

PROF. DR. NICOLAI PETKOV

AND PROF. DR. MANUEL CASTEJÓN LIMAS

and submitted by

LAURA FERNÁNDEZ ROBLES

in fulfillment of the requirements for the Degree of

PHILOSOPHIÆDOCTOR (PH.D.)

León, November 2015

UNIVERSIDAD DE GRONINGEN

INSTITUTO JOHANN BERNOULLI DE MATEMÁTICAS E INFORMÁTICA

UNIVERSIDAD DE LEÓN

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y DE SISTEMAS Y AUTOMÁTICA

TÉCNICAS DE RECONOCIMIENTO DE OBJETOS EN APLICACIONES REALES

Tesis doctoral dirigida por

EL PROF. DR. ENRIQUE ALEGRE GUTIÉRREZ,

EL PROF. DR. NICOLAI PETKOV

Y EL PROF. DR. MANUEL CASTEJÓN LIMAS

y desarrollada por

LAURA FERNÁNDEZ ROBLES

a fin de optar al grado de

DOCTOR POR LA UNIVERSIDAD DE LEÓN

León, Noviembre de 2015

Abstract

This thesis evaluates and proposes object description and retrieval techniques in different real applications. It addresses the classification of boar spermatozoa according to the acrosome integrity using several proposals based on invariant local features. In addition, it provides two new methods for inserts localisation and an automatic solution for the recognition of broken inserts in edge profile milling heads that can be set up on-line without delaying any machining operations. And finally, it evaluates different keypoints clustering configurations for object retrieval and proposes a new descriptor, named colour COSFIRE, in the scope of the *Advisory System Against Sexual Exploitation of Children* project.

Automatic assessment of sperm quality is an important challenge in the veterinary field. In this dissertation, we studied the description of boar spermatozoa acrosomes using image analysis to automatically classify them as intact or damaged. We characterised the acrosomes using invariant local features, particularly SIFT and SURF, improving the results obtained with global texture descriptors. The best results were achieved for the classification of SURF descriptors by k-NN. The overall accuracy was 94.88%, with a higher hit rate in the damaged class, 96.86%, than in the intact one, 92.89%. The opposite behaviour, higher hit rate in the intact class, was yielded by global texture descriptors. In order to overcome the classification of invariant local features with a support vector machines (SVM), we presented an approach which successfully deals with having more than one descriptor per image. Interest points were detected and described using SURF. Our method classifies spermatozoa heads, exploiting that a head usually contains more distinctive points of their own class than doubtful points which could be misclassified. Experiments showed an accuracy of 90.91% (94.94% and 86.87% for the intact and damage classes respectively) which indicates that this approach could be an alternative to consider for classifying invariant local features descriptors. We also proposed an early fusion of invariant local features with global texture descriptors to study the integrity of the head acrosomes, evaluating both SVM with bag of visual words (BoW) and k-NN for

the classification. The concatenation of SURF with Legendre descriptors achieved an accuracy of 95.56% (93.63% in the intact and 97.48% in the damaged class) when classifying using kNN, outperforming the results obtained for both descriptors separately.

Wear evaluation of inserts is a key issue for extending lifetime of cutting tools and ensuring high quality of products. In this thesis, we introduced two image processing methods to automatically localise cutting tools in an edge profile milling head and another one to determine if they are broken. Unlike other machining operations presented in the literature, we were dealing with edge milling head tools for aggressive machining of thick plates (up to 12 centimetres) in a single pass. The studied cutting head tool is characterized by its relatively high number of cutting tools (up to 30) which makes the localisation of inserts a key aspect. We detected the screws that fasten the inserts using a circular Hough transform. In a cropped area surrounding a detected screw, we used Canny's algorithm and a standard Hough transform to localise line segments that characterise insert edges. Considering this information and the geometry of the insert, we identified which of these line segments is the cutting edge. The output of our algorithm is a set of quadrilateral regions around the identified cutting edges that can be used as input to other methods specialised in assessing the state of the cutting edge. Our proposal is very effective (accuracy equals to 99.61%) for the localisation of the cutting edges of inserts in an edge profile milling machine. Following up this result, we studied how to recognise broken inserts because it is critical for a proper tool monitoring system. The method that we presented first localises the screws of the inserts and then determines the expected positions and orientations of the cutting edges using known geometrical information. We computed the distances, called deviations, between the expected cutting edge and the real one to determine if it is broken. We evaluated the proposed method on a new dataset that we created and made publicly available. The obtained results, with a harmonic mean of precision and recall equals to 91.43%, showed that this algorithm is effective and suitable for the recognition of broken inserts in machining head tools. Finally, we proposed a more generic and versatile approach for the localisation of inserts based on trainable COSFIRE filters. It can be automatically configured regardless of the appearance of the inserts. A new function for the computation of the response of the COSFIRE filter was also introduced, outperforming the previous ones. Results, with a harmonic mean of precision and recall equals to 89.89%, improved preceding works based on template matching. Altogether, the results obtained for this application foster further implementation at a working manufacturing environment.

Advisory System Against Sexual Exploitation of Children European project aims to provide a technological solution to help the fight against child pornography. One of

the most challenging tasks in this project was the retrieval of specific objects from collections with a huge amount of images and videos. We evaluated different clustering configurations of SIFT keypoints in relation with their pose parameters: coordinates location, scale and orientation. On the one hand, we used the similarity measure of the closest pairs of keypoint descriptors. On the other hand, we used a Hough transform, with different parametrization values, to identify clusters of at least three points voting for the same pose of an object and we verified the consistency of the pose parameters with the least squares algorithm. Results were computed for a publicly available dataset of 614 images illustrating possible sceneries of a real case. Higher precisions were obtained without clustering at small cuts of the hit list, whereas better precisions were yielded with Lowe's clustering at high cuts. Moreover, colour COSFIRE filters were proposed for the retrieval of colour objects. They add colour description and discrimination power to COSFIRE filters as well as provide invariance to background intensity. Colour COSFIRE filters were presented both for patterns made up of colour lines and for patterns that are colour objects, outperforming standard COSFIRE filters both for retrieval and classification tasks.

The work proposed in this thesis contributes to the understanding and resolution of real applications using object recognition and image classification techniques.

Resumen

Esta tesis evalúa y propone técnicas de descripción y recuperación de objetos en diferentes aplicaciones reales. Aborda la clasificación de espermatozoides de verraco en función de la integridad de los acrosomas utilizando varias propuestas basadas en características locales invariantes. Además, aporta dos métodos nuevos para la localización de plaquitas y una solución automática para el reconocimiento de plaquitas rotas en fresadoras de bordes que puede ser instalada en línea sin retrasar las operaciones de fabricación. Y, finalmente, evalúa diferentes configuraciones de agrupamientos de puntos clave para la recuperación de objetos y propone un nuevo descriptor, denominado color COSFIRE, en el ámbito del proyecto *Advisory System Against Sexual Exploitation of Children*.

La evaluación automática de la calidad del semen es un reto importante en el campo veterinario. En esta tesis, estudiamos la descripción de acrosomas de los espermatozoides de verraco utilizando el análisis de imagen para clasificarlos automáticamente como intactos o dañados. Caracterizamos los acrosomas usando características locales invariantes, en concreto SIFT y SURF, superando los resultados obtenidos con descriptores de textura global. Los mejores resultados se consiguieron en la clasificación de descriptores SURF con k-NN. La tasa de acierto global fue de 94.88 %, siendo superior en la clase dañada, 96.86 %, que en la intacta, 92.89 %. El comportamiento contrario, una mayor tasa de acierto en la clase intacta, se obtuvo con los descriptores globales de textura. Para poder realizar la clasificación de características locales invariantes con máquinas de vector soporte (SVM), presentamos un enfoque que permite, de manera satisfactoria, tener más de un descriptor por imagen. Los puntos de interés se detectaron y describieron mediante SURF. Nuestro método clasifica las cabezas de espermatozoides aprovechando que una cabeza generalmente contiene más puntos distintivos de su propia clase que puntos dudosos que pudieran ser mal clasificados. Los experimentos mostraron una tasa de acierto del 90.91 % (94.94 % y 86.87 % para las clases intacta y dañada respectivamente) lo que indica que este enfoque podría ser una alternativa a considerar en la clasifica-

ción de descriptores de características locales invariantes. También propusimos una fusión temprana de características locales invariantes con descriptores globales de textura para estudiar la integridad de los acrosomas de las cabezas, evaluando la clasificación mediante SVM con un modelo de bolsa de palabras visuales y mediante k-NN. La concatenación de descriptores SURF y Legendre consiguió una tasa de acierto de 95.56 % (93.63 % en la clase intacta y 97.48 % en la dañada) clasificando con kNN, mejorando los resultados obtenidos para ambos descriptores por separado.

La evaluación del desgaste de plaquitas es un aspecto clave para extender la vida de las herramientas de corte y asegurar una alta calidad de las piezas mecanizadas. En esta tesis, introducimos dos métodos de procesamiento de imágenes para localizar automáticamente las plaquitas de un cabezal de una fresadora de bordes y otro para determinar si están rotas. Al contrario que otras operaciones de fabricación presentadas en la literatura, estamos trabajando con cabezales de corte para la manufactura agresiva de gruesas planchas de acero (de hasta 12 centímetros de espesor) en una sola pasada. El cabezal estudiado se caracteriza por su relativamente alto número de plaquitas (hasta 30) lo que hace que la localización de plaquitas sea un aspecto clave. Detectamos los tornillos que sujetan las plaquitas mediante una transformada Hough circular. En una región recortada alrededor de los tornillos detectados, utilizamos el algoritmo de Canny y una transformada Hough estándar para localizar los segmentos que caracterizan los filos de las plaquitas. Considerando esta información y la geometría de las plaquitas, identificamos qué segmentos conforman los filos de corte. La salida de nuestro algoritmo es un conjunto de regiones cuadrangulares alrededor de los filos de corte identificados que pueden ser utilizadas como entrada de otros métodos especializados en la evaluación del estado de los filos de corte. Nuestra propuesta es muy efectiva (la tasa de acierto es igual a 99.61 %) para la localización de filos de corte de plaquitas en cabezales de fresadoras de bordes. Siguiendo este resultado, estudiamos cómo reconocer plaquitas rotas, ya que es una faceta crítica para un sistema de monitorización adecuado. El método que presentamos primero localiza los tornillos de las plaquitas y después determina las posiciones y orientaciones esperadas de los filos de corte utilizando información geométrica. Calculamos las distancias, a las que denominamos desviaciones, entre los filos de corte esperados y reales para determinar si las plaquitas están rotas. Evaluamos el método propuesto en un nuevo conjunto de imágenes que creamos y dimos acceso público. Los resultados obtenidos, con una media armónica de precisión y exhaustividad igual al 91.43 %, mostraron que este algoritmo es efectivo y adecuado para el reconocimiento de plaquitas rotas en cabezales de corte. Finalmente, propusimos un enfoque más genérico y versátil para la localización de plaquitas basado en los filtros entrenables COSFIRE. Puede ser configurado de manera automática independientemente de la apariencia de las plaquitas. También

introducimos una nueva función para el cálculo de la respuesta del filtro COSFIRE, mejorando las anteriores. Los resultados, con una media armónica de precisión y exhaustividad del 89.89 %, superaron los de los trabajos previos basados en la correspondencia de plantillas. En función de todo lo anterior, podemos decir que, los resultados obtenidos para esta aplicación fomentan implementaciones posteriores en un ambiente de manufacturación.

El proyecto europeo *Advisory System Against Sexual Exploitation of Children* tenía como objetivo principal proveer de una solución tecnológica para ayudar en la lucha contra la pornografía infantil. Una de las tareas más desafiantes en este proyecto fue la recuperación de objetos específicos en colecciones que contienen una gran cantidad de imágenes y vídeos. Evaluamos diferentes configuraciones de agrupamiento de puntos clave SIFT en función de sus parámetros de pose: coordinadas de localización, escala y orientación. Por un lado, utilizamos la medida de similitud del par de descriptores de puntos clave más cercanos. Por otro lado, usamos una transformada Hough, con diferentes valores de sus parámetros, para identificar grupos de al menos tres puntos que votaran por la misma pose de un objeto y verificamos la consistencia de los parámetros de pose con un algoritmo de mínimos cuadrados. Los resultados se calcularon para un conjunto de 614 imágenes que dejamos disponible para acceso público y que ilustran posibles escenarios de un caso real. Se obtuvieron precisiones más altas sin agrupamiento para cortes bajos de la lista de imágenes recuperadas, mientras que se consiguieron precisiones más altas con el agrupamiento de Lowe para cortes más altos de dicha lista. Finalmente, se propusieron los filtros de color COSFIRE para la recuperación de objetos de color. Añaden descripción de color y poder de discriminación a los filtros COSFIRE, a la vez que aportan invarianza a la intensidad del fondo. Los filtros de color COSFIRE se presentaron para patrones compuestos por líneas de color y para patrones que son objetos de color, mejorando a los filtros COSFIRE tradicionales tanto en tareas de recuperación como de clasificación.

El trabajo propuesto en esta tesis contribuye al entendimiento y la resolución de aplicaciones reales utilizando técnicas de reconocimiento de objetos y clasificación de imágenes.

Contents

List of Figures

List of Tables

Índice general

Acknowledgements

A doctoral thesis is sometimes portrayed as a solitary endeavour; however the long list that follows absolutely proves the opposite. Many are the people and institutions that have collaborated to allow a successful end to these studies.

Thanks to Junta de Castilla y León for letting me enjoy a grant destined to fund the hiring of recently graduated research personnel; the company TECOI for providing us an edge profile milling head tool and the inserts to create our data set; and Centrotec for providing us the semen samples and for their collaboration in the acquisition of the images.

I am thankful to my supervisors Nicolai Petkov, from University of Groningen, and Enrique Alegre and Manuel Castejón, from University of León, for giving me the opportunity to carry out my doctoral studies under your supervision and guidance. I am deeply grateful for your continuous support, insight and patience. Your invaluable guidance has made possible this thesis work. I also want to thank you for increasing my connections in research and back me for attending conferences and visiting research institutions.

I would like to express my special appreciation and thanks to George Azzopardi for encouraging my research, for your collaboration and good advice, as well as for your friendship. I am also thankful for your invitation to visit the department Intelligent Computer Systems at the Information and Communication Technology faculty of the University of Malta.

I am grateful to professors Alexandru C. Telea, Constantinos S. Pattichis, Kenneth Camilleri and Javier González Jiménez for serving as my assessment committee members at University of Groningen even at hardship. Thanks for your thorough review of this thesis.

A sincere thanks to the administrative stuff for easing my life at University: Janieta de Jong-Schlukebir, Ineke Schelhaas, Desiree Hansen, Esmee Elshof, Ingrid Veltman, Annette Korringa, Barbara Visser, Eloina Panero, María José Rodríguez and Marian Villamediana. The agreement between both Universities would have never

been possible without your help and hard work.

I feel fortunate and honoured to have had the opportunity to share my academic environment surrounded by great scientists and teachers in an enjoyable atmosphere. Thanks Michael Wilkinson, Michael Biehl, Apostolis Ampatzoglou, Paris Avgeriou, Marco Aiello, Doina Bucur, Rocío Alaiz, Camino Fernández, Joaquín Barreiro, Susana Martínez, Anabel Fernández, Chema Foces, Javier Alfonso and many more.

A special thanks to my paranymphs Nicola, the amazing, and Jiapan, the tiger, for agreeing to be by my side during the defence at University of Groningen. I feel very fortunate and happy that the great people of work colleagues (et al.) have become my lovely family in Groningen: Astone, Maestro Ugo, Andreas, Manu, Estefi, Danilo, Walter, Daniel, Renata, Charmaine, Fritz, Ilario, Spyros, Laura, Fatima, Sofia, Mani and Baris, many thanks for your friendship. I am also very grateful to my student colleagues in León because your friendship, presence and advice have helped me during good and hard times of this thesis: Óscar and Maite (*gracias por llevarme al 'zielo'*), Diego, Víctor, Dani, Manu, Miguel, Edu, Guillermo and Claudia.

To my dear friends, both from Spain and Winscho, because in many occasions this work has drift me apart from you. Thanks for your support and friendship.

Words cannot express how grateful I am to my parents for their love, dedication and generosity on raising me up. *Muchas gracias por darlo todo por mí, amarme con todo vuestro corazón y por la educación que me habéis proporcionado. Este logro es mérito vuestro.* To my brother, because our mutual understanding does not stop growing and you have known how to listen to me and give me advice when I needed it most. And in general, to all my family. Sergio and Adrián, you have taught me the most pure definition of love.

Finally, my most heartfelt gratitude to you, Luis. You have had the patience to suffer my absence and difficulties along this trip. Yet, you have always been of immense support. Thanks for your love, affection, empathy and spirit.

Laura Fernández Robles
León
27th November 2015

Chapter 1

Introduction

1.1. Motivation

Object recognition is one of the fundamental tasks in computer vision. It is the process of finding or identifying instances of objects (for example faces, dogs or buildings) in digital images or videos. Object recognition methods frequently use extracted features and learning algorithms to recognise instances of an object or images belonging to an object category. Object class recognition deals with classifying objects into a certain class or category whereas object detection aims at localising a specific object of interest in digital images or videos. Every object or object class has its own particular features that characterise themselves and differentiate them from the rest, helping in the recognition of the same or similar objects in other images or videos. Object recognition is applied in many areas of computer vision, including image retrieval, security, surveillance, automated vehicle parking systems and machine inspection. Significant challenges stay on the field of object recognition. One main concern is about robustness with respect to variation in scale, viewpoint, illumination, non-rigid deformations and imaging conditions. Another current issue is the scaling up to thousands object classes and millions of images, what it is called large scale image retrieval.

In this thesis we particularly address three tasks of object recognition (Dickinson et al., 2009; Li, 2005):

- *Classification*: Given an image patch, decide which of the multiple possible categories is present in that patch.
- *Detection and localisation*: Given a complex image, decide if an specific object of interest is located somewhere in this image, and provide accurate location information on the object.
- *Content-based image retrieval*: provide automated indexing of images for their retrieval from a dataset, according to the detection and localisation of an object of interest.

This dissertation studies some particularities of object recognition through three different applications: classification of boar spermatozoa according to the acrosome integrity; automatic identification of broken inserts in edge profile milling heads and finally retrieval of objects for the *Advisory System Against Sexual Exploitation of Children* project in relation with the evaluation of a number of clustering techniques applied to keypoint descriptors and the improvement of an existing method, COSFIRE filters, by adding the capability of describing objects using also colour information. In the following, the motivation of each application is presented.

1.1.1. Classification of boar spermatozoa according to the acrosome integrity

Better semen quality leads to higher fertilization potential of a sperm sample for artificial insemination, both in medicine and veterinarian fields. Regarding the last one, the assessment of the quality of semen samples is a crucial task for many industries in order to guarantee an optimal product. Specifically, porcine industry aims at obtaining better individuals for human consumption.

In the last years, the Computer-Assisted Semen Analysis (CASA) systems have been applied to the assessment of the seminal quality (Didion, 2008). However, there are three valuable criteria, used by veterinary experts, that these systems do not measure automatically. Those are the number and presence of proximal and distal droplets, the vitality of the sample based on the presence of dead or alive spermatozoa and the integrity of the acrosome membrane. In this work, we deal with the last criteria. Evaluating the state of the acrosomes is important because a higher proportion of spermatozoa with damaged acrosomes causes a lower fertilization potential.

Currently, the evaluation of the acrosome integrity of the spermatozoon heads is carried out visually, using staining techniques and counting the stained spermatozoa. This manual process is subjective to the human observer, time consuming and requires expensive fluorescent microscopes to visualize the stained samples. Industry would benefit from an automatic classification of the acrosome as intact or damaged achieved directly on non stained sperm samples.

This task has been studied using digital images taken on samples without staining and using a phase-contrast microscope. The existing approaches make use of standard texture description of the spermatozoa heads. These solutions need to segment the heads of the spermatozoa, extract the patterns that characterise them and classify those patterns to finally estimate the rate of damaged acrosomes present in the sample (González-Castro et al., 2009). The segmentation itself is a critical task that represents a yet unsolved problem. By using invariant local features (ILF),

this segmentation step can be avoided. In this work we present several approaches where the classification of boar spermatozoa is carried out using different techniques based on ILF.

1.1.2. Localisation of broken inserts in edge profile milling heads

Figure 1.1 shows a milling head that contains indexable cutting tools, also known as inserts. Metallic plater are machined by the turning of the milling head. In this case, each insert has four edges, with the cutting edge being the (nearly) vertical one on the left hand side. In the problem that we present here we have two challenges: the localisation of inserts and their cutting edge; and the identification of broken inserts.

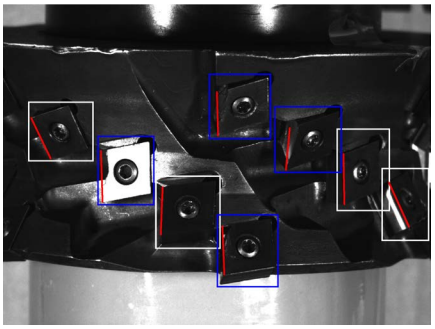
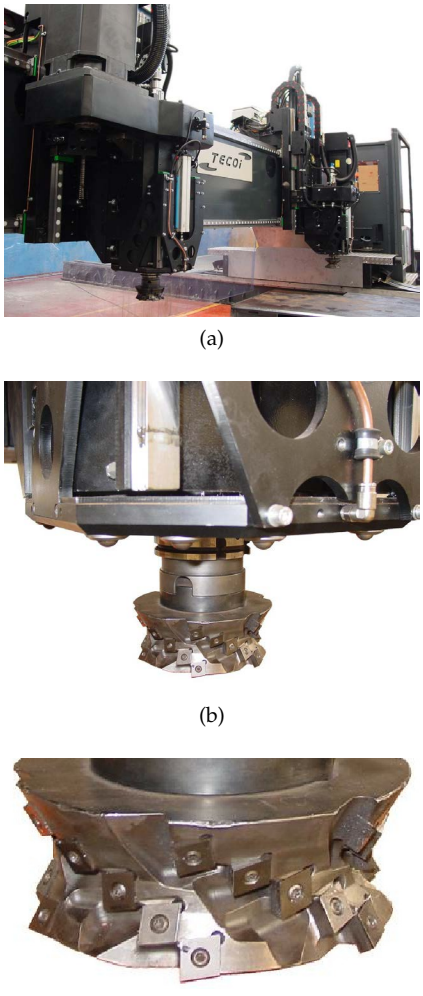


Figure 1.1: Head of an edge profile milling machine. White rectangles mark intact inserts whereas blue rectangles mark broken ones. Red line segments mark the ideal (intact) cutting edges. All markers are provided manually.

Tool wear monitoring (TWM) systems have been widely developed over the last decades for the evaluation of the wear level of cutting tools. The identification of broken cutting tools in a milling machine is an important application as they pose a threat to the stability of a milling head. An unnoticed broken insert may go on working without being detected, and can cause a decay of the quality of the final manufactured product or a breakage of the milling machine itself (Kalvoda and Hwang, 2010).

Figure 1.2 shows a machine which is used to manufacture metal poles of wind towers. Milling is performed in a single pass across very thick and long plates (up to 12 centimetres and 42 meters, respectively) which is not common in standard milling machines. Due to this aggressive operation, part of a cutting edge may be torn out without modifying the external aspect of the remaining part of the insert. The replacement of this broken inserts is quite cheap and requires few time. On the

contrary if a milling head machine collapses, the cost and time for the replacement of the head machine increases heavily.



(a)

(b)

(c)

Figure 1.2: Machine tool for machining of metal poles of wind towers. (a) General view. (b) Detail of the head milling tool. (c) Close-up of the head tool.

As for the localisation of inserts, in our application, the head tool contains 30 rhombohedral inserts leading to 8 to 10 visible inserts per acquired image, which makes the localisation of the inserts a challenging task.

TECOi is a company interested in the development and installation of TWM

systems that are able to automatically detect broken inserts. TECOi provided us such an edge profile milling head tool and the cutting tools to study the automatic inspection of inserts.

1.1.3. Object recognition for content-based image retrieval

Advisory System Against Sexual Exploitation of Children (ASASEC) is a European research project whose goal was to provide a technological solution to help the fight against child pornography. One of the most challenging tasks in this kind of environments consists of retrieving images and videos that contain specific objects from huge datasets. These datasets are collections of many images or videos proven to be related with children exploitation. Finding connections among different scenes or images could help to understand and resolve complex legal cases. In the scope of this project, we have studied the topic of object recognition for content-based image retrieval.

Object recognition for content-based image retrieval (CBIR) aims at retrieving images that contain objects similar to a query object. The retrieved images are sorted in a hit list according to their similarity with the query object. When the object retrieval system is based on query by example, the user chooses an image of interest, also known as query image, and then selects a bounding box in that image, which conforms the region of interest (ROI), containing the query object or object of interest. Then, the ROI is described and the representation of its features is used to match images or videos from a dataset. Changes in pose, scale, orientation, illumination, rigidity, cluttered background or occlusion, among others, make the retrieval of objects a challenging task. Features clustering and object detection become then two crucial tasks which we have partially studied in this thesis.

Invariant local features (ILF) can rely on features clustering in order to improve the matching process. First, the matches between keypoint descriptors of the ROI image and the query image are computed. Then, we should adopt a criterion to assure if there is a real correspondence between images and, if any, the strength of that correspondence. One possibility is to use the distance of the closest match between the ILF descriptors of the ROI and the query image. Thus, the hit list would be created by sorting those computed distances, and a threshold could be set up to decide the minimum value of distance at which a correspondence is considered. However, this could lead to two kinds of errors. On the one hand, the local surroundings of two keypoints could be very similar even when they belong to different objects. On the other hand, unfortunately an ill-selected bounding box makes that the query object comes jointly, partially or completely, with other objects or cluttered background in the ROI. Lowe (2004) suggested to consider clusters of at least 3 features

that agree on an object and its pose for reliable object recognition. He proposed to use Hough transform to identify clusters that vote for the same pose of an object and to perform a geometric verification through least squares solution for consistent pose parameters. Nonetheless, there is a lack of reasoning for the choice of this clustering approach and its theoretical insight. We evaluate both approaches, direct matching to the closest pair of correspondences and the use of Hough transform with least squares verification in the scope of ASASEC project. For the latter, we compare different configurations of clustering sets of keypoints in relation with their pose parameters: coordinates location, scale and orientation obtained with scale invariant feature transform (SIFT) method.

Regarding the object detection, combination of shifted filter responses (COSFIRE) filters have proved to successfully detect given objects in complex scenes. COSFIRE filters are trainable keypoint detection operators that are selective for given local patterns. The approach used with COSFIRE filters is versatile because a filter can be automatically configured for any given prototype pattern, being able to detect identical and similar patterns in digital images. It is inspired by neurophysiological evidence about the visual processing of contour, curvature and shape in the ventral stream of the brain. Therefore, it is also interesting due to the continuing trend of simulating biological vision to design more effective computer vision solutions. Nevertheless, COSFIRE filters have some shortages as for example the inability of dealing with colour digital images. For all the above reasons, we consider that COSFIRE filters can provide a great contribution in recognition and retrieval of colour objects. We add colour description to COSFIRE filters which allows to distinguish objects with similar shape but different colours and to improve object recognition efficiency. Moreover, we also propose a methodology that provides invariance to the background intensity.

1.2. Objectives

The main goal of this dissertation is to select and evaluate appropriate object description and retrieval techniques in different real applications.

Given the previous general goal, we defined the following particular objectives:

1. To evaluate the classification of boar spermatozoa according to the acrosome integrity using approaches based on ILF.

2. To provide an automatic solution for the identification of broken inserts in edge profile milling heads that can be set up on-line without delaying any machining operations.

3. To study two specific fields of object recognition for CBIR in the scope of the *advisory system against sexual exploitation of children* project: the evaluation of different clustering configuration of features and the addition of colour description to COSFIRE filters.

1.3. Main contributions

The main contributions of this dissertation may be summarised as follows:

1. *ILF have been used for the description of the acrosome of boar spermatozoa heads yielding a successful classification of spermatozoa heads as intact or damaged*. The performance of both speeded up robust features (SURF) and SIFT methods has been compared with a number of global texture descriptors (Zernike moments, Haralick features extracted from the original image and from the coefficients of the discrete wavelet transform (DWT), Legendre moments and Laws masks) for the application at hand. SURF has outperformed all the tested global texture descriptors. At the time when this work was published in the form of a conference paper, these were the best results in the literature.

2. *Support vector machine (SVM) has been adapted to deal with several feature vectors per image.* A method to classify SURF features using SVM has been presented and evaluated. This approach can be easily implemented for other ILF and classifiers.

3. *An early fusion of ILF with global texture descriptors has been proposed for the classification of the integrity of the head acrosomes, demonstrating that some of the combinations of global and local features improve the accuracy obtained when using them separately*.

4. *A highly effective and efficient method for the localisation of cutting edges in milling machines has been presented*. Its output is a set of regions surrounding cutting edges, which can be used as input to other methods that perform quality assessment of the edges. It is based on circular Hough transform to find the screws that fasten the inserts and edge detection and standard Hough transform to localise the cutting edge.

5. *A novel method has been introduced for the effective description and classification of inserts, as broken or unbroken, with respect to the state of their cutting edges*. It computes the gradient magnitudes and the deviations of the real cutting edges from the ideal ones in order to classify the inserts of a milling head tool. The

time that this method requires for the inspection of the head tool is below the resting time of the machine.

6. *Another, more versatile and generic, method for the localisation of inserts has been presented*. It differs from the previous one in the way that it considers independently each image. It is based on COSFIRE filters and it can be automatically configured regardless of the appearance of the inserts. A new metric for the computation of the response of the COSFIRE filter has been introduced, outperforming the previous ones. It has obtained better results than preceding works based on template matching.

7. *Different clustering configurations of SIFT keypoints in relation with their pose parameters: coordinates location, scale and orientation have been evaluated*. On the one hand, the similarity measure of the closest pairs of keypoint descriptors has been used. On the other hand, we have used a Hough transform, with different parametrization values, to identify clusters of at least three points voting for the same pose of an object and we have verified the consistency of the pose parameters with the least squares algorithm.

8. *Colour COSFIRE filters have been proposed, adding colour description and discrimination power to COSFIRE filters as well as providing invariance to background intensity*. Colour COSFIRE filters have been presented both for patterns made up of colour lines and for patterns that are colour objects. It has outperformed results for CBIR and classification tasks on COIL data with respect to standard COSFIRE filters.

1.4. Thesis Organization

In this section the structure of this doctoral thesis is described. This first introductory chapter has been focused on motivating the work presented in this dissertation, its main objectives and original contributions. Now, the remaining chapters of this thesis are organised as follows.

Chapter 2 contains a review of object recognition methods as well as a more specific review of the state of the art for each studied application. Thus, it comments published methods that deal with description and classification of boar spermatozoa in relation with the state of the acrosome heads. Then, it studies the literature research that evaluates tool wear monitoring systems and specifically how they relate with the localisation of cutting tools and the identification of broken inserts. And finally, it reviews object recognition methods for CBIR focusing on Hough transform and COSFIRE filters for object recognition.

Chapter 3 addresses the classification of boar spermatozoa according to the acrosome integrity using approaches based on ILF. A comparison of SIFT and SURF methods against some global texture descriptors in a quite large dataset is shown in this chapter. SVM algorithm is adapted to deal with several feature vectors per image in order to classify SURF descriptors. This chapter also introduces an early fusion of ILF with global texture descriptors for the description of the spermatozoa heads.

Chapter 4 presents an automatic solution for the identification of broken inserts in edge profile milling heads that can be set up on-line without delaying any machining operations. Together with it, two methods for the localisation of inserts are proposed in this chapter. One based on Hough transform and edge detection that solves the specific problem at hand and whose output is a set of regions surrounding cutting edges. This output can be used as input to other methods that perform quality assessment of the edges. And a second one based on COSFIRE filters (Azzopardi and Petkov, 2013c) that can be automatically configured regardless of the appearance of the inserts. This chapter also introduces a new metric for the computation of the response of the COSFIRE filter.

Chapter 5 studies two specific fields of object recognition for CBIR in the scope of ASASEC project. Firstly, different clustering configurations of SIFT keypoints in relation with their pose parameters: coordinates location, scale and orientation are evaluated. Secondly, this chapter presents colour COSFIRE filters that add colour description and discrimination power to COSFIRE filters (Azzopardi and Petkov, 2013c) as well as provide invariance to background intensity.

Chapter 6 contains a summary with the conclusions of this thesis and gives an outlook of possible future work lines to extend the presented work.

Regulations about the Ph.D. studies at the University of León claim that if a doctoral thesis is not written in Spanish, at least the table of contents, conclusions, and a résumé of each chapter must be written in Spanish. In order to comply with these regulations, we include a translation of the conclusions in Chapter 7, and a summary of all chapters in Part II.

Chapter 2

State of the art

In the last decades, there has been substantial work in the computer vision field that tackles the problem of object recognition. Here we present a brief survey of different approaches on object recognition.

Some reviews divide object recognition approaches in three categories. *Model-based* methods deal with the representation and identification of a known three dimensional (3-D) objects (boxes, spheres, cylinders, cones, surface of revolution, etc.). Similarly, *shape-based* methods represent an object by its shape and/or contour. In contrast, *appearance-based* models use the appearance of the object usually under several two dimensional (2-D) views.

Another way of classifying object recognition techniques distinguishes between *local* and *global* approaches. *Local* methods search for salient regions or points that characterise the object of interest such as corners, edges or entropy. Later, these regions are typified by given descriptors. The local descriptors of the object of interest and the local descriptors of the test image are then compared for object recognition purposes. In contrast to that, *global* methods model the information content of the whole object of interest. This information can come from simple statistical measures (such as mean values or histograms of features) to more advanced dimensionality reduction techniques. *Global* methods allow to reconstruct the original image providing robutness to some extend whereas *local* methods can better cope with partly occluded objects.

Local appearance based object recognition methods need to detect and describe distinctive regions or keypoints in an image. As for the detection, we can differentiate *corner based* detectors, *region based* detectors and *others*.

Corner based detectors locate keypoints and regions which contain a lot of image structure like edges. Corners can be defined as points with low self-similarity in all directions. The self-similarity of an image patch can be measured by taking the sum of squared differences (SSD) between an image patch and a shifted version of itself. The most popular corner based detector is the one of Harris and Stephens (1988). It works by computing a response function across all the image pixels. Then, those exceeding a threshold, also known as locally maximal, are considered corners. The response function is obtained from the Harris matrix computed from

image derivatives. The Harris-point detector achieves a large number of keypoints with sufficient repeatability (Schmid et al., 2000). The main advantage of this detector is the high computation speed whereas the main disadvantage is that the detector is only invariant to rotation since no information about scale and orientation is provided. Harris-Laplace detector add invariance to scale and is based on the work of Lindeberg (1998) that studies the properties of scale space. Mikolajczyk and Schmid (2002) proposed Harris-Affine detector by extending the Harris-Laplace detector in order to achieve invariance to affine transformations. It is based on the shape estimation properties of the second moment matrix. The main disadvantage of the Harris-Affine detector is the increase in time computation.

Region based detectors locate local blobs of uniform brightness and are therefore suited for uniform regions or regions with smooth transitions. Hessian matrix detectors (Mikolajczyk et al., 2005) are similar to Harris detectors. The Hessian matrix is computed from the second image derivatives, thus, this detector responds to blobs-like structures. Keypoints are selected based on the determinant of the Hessian matrix after non-maximum suppression. The main drawback is that it provides only rotational invariance properties. Similarly, Hessian-Laplace detectors have scale invariance properties and Hessian-Affine detectors add invariance to affine transformations, (Mikolajczyk and Schmid, 2002). Instead of a scale normalised Laplacian, Lowe (1999, 2004) uses an approximation of the Laplacian, namely difference of Gaussian function (DoG), by calculating differences of Gaussian blurred images at different adjacent local scales. The main advantage is the invariance to scale but it is punished by an increase in runtime. Maximally stable extremal regions (MSER) (Matas et al., 2004) are regions that are either darker or brighter than their surroundings and that are stable across a range of thresholds of the intensity function. If single keypoints are needed, they are usually consider as the centres of gravity of each MSERs. The number of MSERs detected is rather small in comparison with previous mentioned detectors, but Mikolajczyk et al. (2005) affirms that the repeatability is higher in most cases.

An example of *other* approaches different to corner or region detectors can be the entropy based salient regions detectors (Kadir et al., 2003; Kadir and Brady, 2001, 2003). They consider the grey value entropy of a circular region in the image in order to estimate the visual saliency of a region. The main drawback is that it is time consuming, specially for the affine invariant implementation (Kadir et al., 2004). Tuytelaars and Gool (1999) and Tuytelaars and Van Gool (2004) proposed two detectors, intensity based regions (IBR) and edge based regions (EBR). Some works locally describe the whole object in a dense way such as bag of words (BoW) description based on dense SIFT and histogram of oriented gradients (HOG). BoW (Sivic and Zisserman, 2009) is a vector of occurrence counts of a vocabulary of local image

features. HOG (Dalal and Triggs, 2005a) counts occurrences of gradient orientation in a dense grid of uniformly spaced cells and uses overlapping local contrast normalization to improve accuracy.

After region or point detection, features descriptors should be computed for describing the regions or the local neighbourhoods of the points respectively. One can distinguish among *distribution based* descriptors, *filter based* descriptors and *other* methods.

Distribution based descriptors represent some properties of given regions by histograms. Usually these properties come from the geometric information of the points and local orientation information in the region. Probably the most popular descriptor is scale invariant feature transform (SIFT), developed by Lowe (1999, 2004). Actually, he proposed a combination of both SIFT detector and descriptor, where SIFT detector is the DoG previously discussed. For obtaining SIFT descriptors, the local image gradients are measured at the selected scale in the region around each keypoint. These are usually transformed into a representation of a 4×4 array of histograms with 8 orientation bins in each, leading to a 128 element feature vector for each keypoint. SIFT is invariant to uniform scaling and orientation and partially invariant to affine distortion and illumination changes and allows for object recognition under clutter and partial occlusion. The main disadvantage of SIFT is the high computational time required. Many versions of SIFT have been proposed. Ke and Sukthankar (2004) reduced the dimensionality of the SIFT descriptor by applying principal component analysis (PCA) to the scale-normalised gradient patches instead of gradient histograms on the keypoints. Gradient location-orientation histograms (GLOH) (Mikolajczyk et al., 2005) try to obtain higher robustness and distinctiveness than SIFT descriptors. Authors divided the keypoint patch into a radial and angular grid leading to a higher dimensional descriptor that is reduced by applying PCA to the 128 largest eigenvalues. Spin images (Johnson and Hebert, 1999) for 2-D images (Lazebnik et al., 2003) use a 2-D histogram of intensity values and their distance from the centre of the region. Each row of the 2-D descriptor represents the histogram of the grey values in an annulus distance from the centre. This descriptor is invariant to in-plane rotations. Belongie et al. (2002) introduced shape context descriptors that compute a descriptor by the distribution of relative point positions and the corresponding orientations collected in a histogram. It is though more sensitive to the positions nearby the keypoints. In other line of research, locally binary patterns (LBP) introduced by Ojala et al. (1996) is a texture descriptor based on a simple binary coding of thresholded intensity values. It has been extended in many directions and used for diverse applications with good performance.

Filter based descriptors capture the properties of the regions or patches around a keypoint by means of filters. Differential-invariant descriptors (Koenderink and van

Doorn, 1987; Schmid and Mohr, 1997) are sets of differential operators obtained by combination of local derivatives. The main drawback is that they are only invariant to rotation. Steerable filters are obtained by a linear combination of some basis filters that yield the same result as the oriented filter rotated to a certain angle, for example in (Freeman and Adelson, 1991). Complex filters encompasses filters with complex valued coefficients (Baumberg, 2000; Carneiro and Jepson, 2003; Schaffalitzky and Zisserman, 2002).

Moment invariants are one example of *other* methods that aim at the description of local regions. For instance, Van Gool et al. (1996) introduced generalized intensity and color moments to use the intensity or multispectral nature of image data for image patch description whereas Mikolajczyk et al. (2005) presented gradient moments.

Global appearance-based methods for object recognition directly describe the whole object patch in the image. Traditional methods for texture, shape, colour, etc. description can be used as well for object recognition purposes. Some other methods project the object input images into a lower dimensional subspace, hence they are called *subspace* methods. Among them, the most representative ones could be PCA (Jolliffe, 2002) that was introduced to by Kirby and Sirovich (1990) in the field of computer vision and became popular when Turk and Pentland (1991) successfully applied it for face recognition. Non-negative matrix factorization (NMF) (Paatero and Tapper, 1994; Shen and Israël, 1989) was introduced by Lee and Seung (1999) for object representation tasks. Independent component analysis (ICA) (Ans et al., 1985; Hyvärinen and Oja, 2000) became widely known when it was used in signal processing for separation of mixed audio signals (Comon, 1994; Jutten and Herault, 1991). Bartlett et al. (2002) and Draper et al. (2003) proposed two approaches based on ICA for object recognition purposes. Canonical correlation analysis (CCA) (Hotelling, 1936) aims at finding pairs of directions that yield the maximum correlation between two random variables.

Once the description of objects is performed, we need to recognise if they correspond with the object or class of interest. Different techniques have been proposed for comparing the descriptors of both the object of interest and an input image. Ones rely on computing similarities or distance measures among them. There are many in the literature, to name some: Kullback-Leibler divergence, Hellinger distance, total variation distance, Rényi's divergence, Jensen-Shannon divergence, Lévy-Prokhorov metric, Bhattacharyya distance, earth mover's distance, energy distance, signal-to-noise ratio distance, Mahalanobis distance, Minkowski distance, distance correlation, Fisher information metric or cosine similarity. Whereas, image classifiers are algorithms for the classification of the image descriptors into classes, as for example Fisher's linear discriminant, logistic regression, Naive Bayes classi-

fier, perceptron, support vector machines, Kernel estimation, k-nearest neighbours, boosting, decision trees, random forests, neural networks or learning vector quantization.

This was just a short review of some very remarkable works in object recognition but many more have been proposed and applied during the last decades. However, it is not the purpose of this thesis to go further in detail. For more information about the topic we refer the reader to (Andreopoulos and Tsotsos, 2013; Roth and Winter, 2008; Matas et al., 2004). In the following, we provide a review of the state of the art for each studied application.

2.1. Classification of boar spermatozoa according to the acrosome integrity using ILF

Works dealing with the classification of the acrosome integrity of boar spermatozoa are mainly based on texture description. Following a time line of recent works, González et al. (2007) used first order statistics and Haar features in combination with Wavelet coefficients. The classification using neural networks (NN) on a dataset of 363 instances extracted from images of 2560×1920 pixels reached a hit rate of 92.19%. Later on, with the same dataset, Alegre et al. (2008) computed the gradient magnitude along the outer contours of the sperm heads and classified with a learning vector quantization (LVQ) of four prototypes obtaining a hit rate of 93.2%. For 393 similar instances Alegre et al. (2009) compared Haralick features, Laws masks, Legendre and Zernike moments classified with supervised (k-nearest neighbours, k-NN, and neural networks) and unsupervised (linear discriminant analysis, LDA, and quadratic discriminant analysis, QDA). Haralick features and LDA achieved the best hit rate of 93.89%.

Furthermore, Alegre et al. (2012) characterized acrosomes by means of the first order statistics derived from the gray level co-occurrence matrix (GLCM) of the image, both computed from the original image and from the coefficients yielded by the discrete Wavelet transform (DWT). Experimental results in a dataset of 800 instances coming from images of 780×580 pixels reached a hit rate of 94.93% with a multilayer perceptron classifier. It outperformed moments-based descriptors (Hu, Zernike and Legendre) and k-NN classifiers. Alegre et al. (2013) computed a local texture descriptor for each point in seven inner contours. They classify using relevance LVQ, obtaining a hit rate of 99%. Experiments were tested in only 360 instances coming form images of 2560×1920 pixels.

To the best of our knowledge, the last work on this field, by García-Olalla et al. (2015), combined local and global texture descriptors and contour descriptors.

Global texture description was obtained from the GLCM of the original image and the four sub-images of first level of decomposition with the DWT based on Haar wavelets. LBP and Fourier shape descriptors provided the local texture and the contour descriptions, respectively. They performed an early fusion by concatenation of the descriptors and the 10-fold classification using Support Vector Machine backed by a least squares training algorithm and a linear kernel yielded an accuracy of 99.19%. A total of 1851 instances coming from images of 780×580 pixels conformed the dataset. To our knowledge, there is no work that focus on the integrity evaluation of sperm cells of humans or any animal using an invariant local features approach based on the detection and description of keypoints.

Recent object description approaches rely on local features rather than in global descriptors since local description can reliably detect highly distinctive keypoints of an image. Therefore, a single feature can be correctly matched with high probability against a large database of features, providing a basis for object and scene recognition. Another advantage is the capability for still recognising partly occluded objects.

The development of image matching by using a set of local interest points was definitively efficient when Lowe (1999) presented SIFT, introducing invariance to the local feature approach. Since then, the computer vision community has been very active presenting many improvements based on SIFT method. Although over a decade old, Lowe's algorithm has proven very successful in a number of applications using visual features and mainly object recognition. The main problem associated with it has been the large computational burden imposed by its high dimensional feature vector what, in recent years, led to the emergence of new proposals mainly focused on obtaining equally robust descriptors but more computationally efficient. The first of a series was FAST (Rosten and Drummond, 2006) which uses a machine learning approach to derive a feature detector similar to Harris, SUSAN (Smith and Brady, 1995) or DoG but much faster that them, but with the disadvantage that is not very robust to the presence of noise. Later, Bay et al. (2008) introduced SURF, outperforming previously proposed schemes with respect to repeatability, distinctiveness, robustness and speed. Recently, Ozuysal et al. (2010) showed a fast keypoint recognition method using random Ferns which avoids the computationally expensive patch preprocessing by using hundreds of simple binary features. Following this idea and due to the need of running vision algorithms on mobile devices with low computing power and memory capacity new approaches have been developed.

A step farther was proposed by Calonder et al. (2010). They use binary strings as a feature point descriptor, called BRIEF that is highly discriminative even when using few bits and can be computed using simple intensity difference test. Another big

advantage of BRIEF is that the matching can be performed by using the Hamming distance, which is more efficient to compute than the Euclidean distance employed in most of the invariant local features detectors. Both aspects convert BRIEF in a faster descriptor in construction and matching but as it is not invariant to large in-plane rotations, it is not suitable for some object recognition tasks.

Another faster than the classical SIFT and SURF but at comparable matching performance is the BRISK detector (Leutenegger et al., 2011). BRISK relies on a configurable circular sampling pattern from which it computes brightness comparison to form a binary descriptor string. Its detection methodology is inspired in the adaptive corner detector proposed by Mair (Mair et al., 2010) for detecting regions of interest in the image, which they named as AGAST. It is essentially an extension of FAST (Rosten and Drummond, 2006), proven to be a very efficient basis for feature extraction. With the aim of achieving invariance to scale BRISK goes a step further by searching for maxima not only in the image plane, but also in scale-space using the FAST score as a measure for saliency.

An evolution of the above-mentioned methods is the ORB descriptor (Rublee et al., 2011) that builds its proposed feature on FAST and BRIEF, standing its name for Oriented FAST and Rotated BRIEF (ORB). Their authors address several limitations of these techniques, mainly the lack of rotational invariance present in BRIEF. They add a fast and accurate orientation component to FAST and, at the same time, they present an efficient way to compute the oriented BRIEF features. Furthermore, the ORB descriptor use a learning method for de-correlate BRIEF features under rotational invariance, leading to better performance in nearest-neighbour applications. ORB was evaluated using two datasets: image with synthetic in-plane rotation and added Gaussian noise, and a real-world dataset of textured planar images captured from different viewpoints. As their authors pointed out, ORB outperforms SIFT and SURF on the real-world dataset, both the outdoor and the indoor one, what makes this method a good choice for object recognition.

Despite the new methods, SIFT and SURF are still the most popular methods for object recognition in present applications.

2.2. Localisation of broken inserts in edge profile milling heads

Tool wear monitoring (TWM) systems have been widely developed over the last decades for the evaluation of the wear level of cutting tools. The current state of the art in TWM presents two approaches known as direct and indirect methods. Indirect techniques can be applied while the machine is in operation. These methods

evaluate the state of the inserts through variables (e.g. cutting forces and vibrations) that are typically affected by noisy signals (Jurkovic et al., 2005; Kurada and Bradley, 1997; Wang et al., 2005; Zhang and Zhang, 2013). On the contrary, direct techniques monitor the state of the cutting tools directly at the cutting edge when the head tool is in a resting position, Pfeifer and Wiegers (2000). As to direct methods, image processing and computer vision techniques are the most popular tools for measuring flank and crater wear, Zhang and Zhang (2013). Ongoing progress in the fields of machine vision, computing hardware and non-tactile applications has permitted the implementation of reliable on-line TWM systems, Dutta et al. (2013). The methods that we present in this thesis fall into the direct approach category.

There is a large body of work in the literature that evaluates the state of given inserts without having to localize them in images. Castejón et al. (2007) disassembled the inserts and located them in a tool-fixture after machining a full pass along the part which allowed to keep constant the flank location in the image. Lim and Ratnam (2012) placed the inserts in a scanner. Xiong et al. (2011) took out the tool and placed the inserts on a plate. Other methods captured images directly on the head tool but they are focused on ball-end milling cutters (Zhang and Zhang, 2013) or microdrills, in which only two flutes and therefore two cutting tools are present. For the latter, some works developed their own acquisition system, Su et al. (2006) to capture the microdrill from the top and Kim et al. (2002) from the side.

Another papers deal with face milling heads where it is easy to set the acquisition system to only capture one insert per acquired image. Jurkovic et al. (2005) made use of a halogen lamp, LED illumination and a telecentric lens system capturing one insert at a time. Wang et al. (2006) only used a fibre-optic lighting to illuminate the insert. Pfeifer and Wiegers (2000) captured the same insert under several lighting positions to combine contour information of all images. Whereas Sortino (2003) presented a portable acquisition system. In our application, the head tool contains 30 rhombohedral inserts leading to 8 to 10 visible inserts per image, which makes the localisation of the inserts a new and challenging task.

In relation with the recognition of broken inserts, approaches based on texture analysis have been widely used in the literature for wear monitoring when dealing with machining operations (Dutta et al., 2013). The GLCM texture features obtained from images or subimages of the cutting inserts have been used to evaluate tool wear in turning processes. Danesh and Khalili (2015) combined GLCM features with undecimated wavelet decomposition to describe the state of the cutting tool. Kerr et al. (2006) computed five Haralick features extracted from the GLCM. Barreiro et al. (2008) estimated three wear levels (low, medium, high) of the tool insert by means of the Hu and Legendre invariant moments. Datta et al. (2013) used two texture features based on Voronoi tessellation to describe the amount of flank wear

of a turning tool. Prasad and Ramamoorthy (2001) captured a pair of stereo images and used the sequential similarity detection algorithm for pattern matching to measure crater wear depth with a back propagation neural network. However, the machines that we are dealing with in this study use an aggressive edge milling in a single pass to mechanise thick plates. This may cause the breakage of cutting tools, such as the examples shown in Fig. 1.1. Part of a cutting edge may be torn without harming the texture of the remaining part of the insert. For this reason we believe that texture and depth features, as well as ILF, are not suitable for the concerned application.

Other approaches use the contours of the cutting tools to determine the state of the inserts. For instance, Atli et al. (2006) classified drilling tools as sharp or dull using a new measure namely DEFROL (deviation from linearity) to the Canny-detected edges. Makki et al. (2009) captured images of a drilling tool at 100 rpm rotation speed and used edge detection and segmentation methods to describe the tool wear as the deviation of the lip portion. Also, Chethan et al. (2014) compared image areas of the tool obtained through a texture-based segementation method before and after cutting in order to determine the state of a drilling tool. For turning operations, Shahabi and Ratnam (2009) applied thresholding and subtraction of worn and unworn tool images to measure the nose worn regions.

Some papers also deal with micro milling or end milling in this line of work. Otieno et al. (2006) compared images captured before and after the usage of two fluted micro and end mills thresholded by an XOR operator. Neither edge detection, nor tool wear quantification and nor wear classification was performed. Zhang and Zhang (2013) also compared images of ball-end milling cutters before and after machining process in order to monitor the state of the tool. Liang et al. (2005) presented a method based on image registration and mutual information to recognize the change of nose radius of TiN-coated, TiCN-coated and TiAlN-coated carbide milling inserts for progressive milling operation. They perform logic subtraction of two images before and after milling. The mentioned works share one common requirement; they all must have an image of the intact tool to evaluate any discrepancies of a new image of the same tool.

We propose a novel algorithm that evaluates the state of cutting edges without requiring image references of intact cutting tools. This avoids calibrating the system each time an insert is replaced and allows to free memory after each monitoring. It automatically determines the ideal position and orientation of the cutting edges in a given image and computes the deviation from the real cutting edges. This means that from a single image we can determine the broken and unbroken inserts.

2.3. Object recognition for content-based image retrieval: Hough transform and COSFIRE filters for object recognition

Content-based image retrieval (CBIR) is a technique for retrieving images from a collection on the basis of features, such as colour, texture and shape, that can be automatically extracted from the images. We deal with the description of an object of interest in order to retrieve images in an image collection that present regions with high similarity with respect to the object. Typically, such a CBIR system describe both the object of interest and the images of the dataset using object description techniques. These techniques has been previously reviewed in this chapter. Then, a similarity or distance measure is computed and sufficiently similar images (over a given threshold) are retrieved in an indexed hit list. The hit list is sorted from more to less similarity or from less to more distance if similarity or distance measures were used respectively. Sometimes a relevance feedback approach is applied afterwards. Relevance feedback aims at retrieving a new hit list by using information about the relevance of the previously retrieved results. A recent review of relevant feedback systems for CBIR is performed in (Yasmin et al., 2014).

Object recognition for CBIR basically consists of two difficult tasks: identifying objects on images and fast searching through large collections of identified objects, since usually CBIR task deals with large datasets. Identifying objects on images is a challenge because the same objects and scenes can be viewed under different imaging conditions. There are many works dedicated to object recognition for CBIR. Some of them are based on textures (Chang and Kuo, 1993), (Francos et al., 1993), (Jain and Farrokhnia, 1990) and (Smietanski et al., 2010), shape (Jagadish, 1991), (Kauppinen et al., 1995) and (Veltkamp and Hagedoorn, 2001), color representation (Huang et al., 1997), (Kiranyaz et al., 2010) and (Pass and Zabih, 1996) or edge detectors (Ogiela and Tadeusiewicz, 2002), (Ogiela and Tadeusiewicz, 2005) and (Zitnick and Dollár, 2014). Recently local invariant features have gained a wide popularity (Lowe, 2004), (Matas and Obdrzalek, 2004), (Mikolajczyk and Schmid, 2004), (Nister and Stewenius, 2006) and (Sivic and Zisserman, 2003). As for the second task, to find similar images to a query image, we need to compare all feature descriptors of all images with the feature descriptors of the object of interest usually by some distance measures. Such comparison is highly time consuming and the literature presents many methods based on some form of approximate search, for instance using a voting scheme, histograms of clustered keypoints or building hierarchical trees. The scope of this dissertation does not concern this task.

2.3.1. Model fitting and Hough transform for object recognition

Robust model building, is responsible for providing a prediction model that works well not only with the training data but also with new data from fresh images not considered on the determination of the parameters of the model. As different behaviours are to occur in real case studies, cluster analysis techniques will be of much help on isolating different populations for better determining their characteristics. The use of those varies on the model fitting technique applied later on, but in general terms, having acquired such degree of knowledge about their distribution empowers the successive analyses for more accurate results. Classification analysis techniques can be divided in two large groups: supervised classification -also known as discriminant analysis- and non-supervised classification -also known as clustering-.

The purpose of cluster analysis (Gordon, 1999) is to identify groups of differentiated behavior within the samples obtained as they are supposed to have been generated by different populations or different states of the generating process. That could be a general definition that in the context of computer vision can be translated to identify different groups of elements pertaining to different objects, images, textures, etc. Predominant clustering techniques can be classified mainly in hierarchical techniques and partitioning techniques.

The Hough transform (Hough, 1962) is a popular technique in the computer vision area. Amongst its numerous advantages robustness can be highlighted. Moreover, it is a fairly efficient algorithm suitable for situations like the one studied in this dissertation where a large set of pictures might be involved (Dattner, 2009). Though originally defined for identifying simple shapes in images, like lines and circles, its use has been extended to more general shapes, allowing for detection of multiple instances of an object that might even be partially occluded. Ballard (1981) introduced the generalised Hough transform (GHT) that modifies the Hough transform using the principle of template matching. In this way, the Hough transform can be used to detect any object described by its model, not only by an analytic function. The complexity relies on the number of parameters chosen to represent the complex shape, with search times increasing in exponential order. Fortunately, the Hough transform can be easily implemented on parallel computing systems as each image point can be treated independently (Illingworth and Kittler, 1988) using more than one processing unit.

Essentially, the Hough transform converts sets of points in an image to a parameter space. Thus, two points can be represented in the parameter space as a point in a two dimensional space whose axes represent the two parameters needed to define the line over those two original points in the image. Similarly, the points of

circles, ellipses and parabolas in the image can be transformed to points in a new space of parameters. Shape parameterization extends this idea further by means of high-dimensional parameterization that can be decomposed into smaller sets of parameter to determine sequentially.

One of the complexities that appear when using the Hough transform is that for a single image, many points can be chosen to belong to a single line, and thus many lines can be adjusted with the whole dataset. Model fitting comes to the rescue in order to choose the best model to use. Additionally, by using the Hough transform a voting scheme can be adopted (Illingworth and Kittler, 1988) which is one of the most common manners of applying the algorithm.

Hough transform has found immense practical applications in object recognition tasks. Lowe (2004) used Hough transform to identify clusters of features descriptors that belong to a single object by using each feature to vote for all object poses that are consistent with the feature. Tang et al. (2015) proposed a novel object detection approach based on Hough transform. They introduced a multi-scale voting scheme in which multiple Hough images corresponding to multiple object scales can be obtained simultaneously to efficiently handle object scale changes. For 3-D object detection, Silberberg et al. (1984) described an iterative Hough procedure with an initial sparse and regular subset of parameters and transformations that is evaluated for goodness-of-fit. Then the procedure is repeated by successively subdividing the parameter space near current best estimates or peaks. A Hough voting approach was used by Tombari and Di Stefano (2010) for object detection in 3D scenes with significant occlusion and clutter. Tong and Kamata (2010) used a 3D Hough transform to obtain a spectrum on which 3D features are concentrated on the sphere and apply Hilbert scanning on the sphere to match the objects. They affirmed to be able to match the object of interest even in overlapping and noise situations.

Medical images also uses Hough transform for automatic object recognition purposes. Golemati et al. (2005) efficiently used Hough transform to automatically segment healthy arterial wall lumen from B-mode ultrasound images of the carotid artery. McManigle et al. (2012) proposed a two-step Hough transform to find an annular approximation of the left ventricular myocardium in short-axis echo slices. Ecabert et al. (2008) used a 3-D implementation of the generalised Hough transform to localise the heart in images obtained with tomography scanners and Zhang et al. (2010) used it for the localisation of livers in computed tomography scans. Brummer (1991) used 3-D Hough transform for automatic detection of the longitudinal fissure in tomographic scans of the brain. Guan and Yan (2011) used Hough transform for blood cell segmentation and Zhang et al. (2012) for estimating 3D orientation of a vertebra. Hough transform matched the projections of the standard 3D primitive with the vertebral contours in biplanar radiographs, where the projected

contours were generated from the 3D model by sampling the viewing sphere with a hierarchical scheme. Tino et al. (2011) showed that a probabilistic model-based Hough Transform (HT) applied to the hexaMplot can be used to detect groups of coexpressed genes in the normal-disease-drug samples.

2.3.2. COSFIRE filters for object recognition

COSFIRE filters were introduced by Azzopardi and Petkov (2013c) for the localisation of given local patterns that consist of combinations of contour segments. They have proved to be highly efficient for applications based on patterns made up of lines such as vessel delineation (Azzopardi et al., 2015; Strisciuglio et al., 2015), vascular bifurcations (Azzopardi and Petkov, 2013a), hand written digits recognition (Azzopardi and Petkov, 2013b) or the differentiation of typical line patterns in skin disease (epidermolysis bullosa acquisita) (Shi et al., 2015).

Their effectiveness for detecting objects has also been demonstrated. Azzopardi and Petkov (2013c) applied COSFIRE filters for the recognition of three types of traffic signs and achieved perfect detection and recognition performance for a dataset of 48 traffic scenes. A trainable hierarchical object recognition model for the application of a home tidying pickup robot was presented in (Azzopardi and Petkov, 2014) allowing to detect deformable objects embedded in complex scenes without prior segmentation. In this case, a dataset of 60 images for shoes detection was used with perfect detection and recognition performance. Guo et al. (2015) introduced inhibition to COSFIRE filters. A COSFIRE filter responds to a pattern made up by a combination of contour segments presented in the configuration. However, it will also respond to patterns that contain a combination of the previous segments together with other contour segments. Guo et al. (2015) overcame this issue by subtracting a fraction of the combined responses of inhibitory part detectors from the combined responses of excitatory part detectors. Both excitatory and inhibitory parts are automatically determined. They applied this new filters for the recognition of architectural and electrical symbols demonstrating the effectiveness of the method even for noisy images. Nevertheless up to our knowledge, there are no works that use colour information to improve the detection of colour objects neither that provides an automatic solution for the invariance to background intensity, that are two of the contributions presented in this dissertation.

Chapter 3

Classification of boar spermatozoa according to the acrosome integrity

3.1. dataset

A digital camera Basler Scout scA780-54fc was connected to a computer with specific software to control the camera and to an epifluorescence microscope Nikon E-600. This microscope allows to visualise both phase contrast and fluorescence images of the samples. For each semen sample, first the visible light filter DIA-ILL was placed to observe the sample in positive phase contrast, the lens was set up in focus and an image was captured with the visible light power supply on. Right afterwards, the fluorescence filter B-2A EX 450-490, DM 505, BA 520 was placed and another image was taken with the visible light turned off and the fluorescence light on. Therefore, each sample produced two images. Phase contrast images, in gray scale, were used to test our experiments whereas fluorescence images, in colour scale, were used to create the ground truth. Under fluorescence, the spermatozoa with damaged acrosome reacts and fluoresces bright green while the spermatozoa with intact acrosome does not react neither fluoresce due to the preparation sample explained by Sánchez et al. (2006), Fig. 3.1. All images have been acquired at CENTROTEC, an Artificial Insemination Centre which is a spin-off of the University of León, under the guidance of veterinary experts. Semen samples come from boars of three different breeds: Piyorker, Large White and Landrace.

Images were taken with a resolution of 780×580 pixels with $100\times$ magnification of the microscope. Thus, usually no more than three or four heads have been acquired per snapshot. Since most of the spermatozoa come from different takings, illumination is not completely constant.

For each image, the heads of the spermatozoa were cropped. Overlapped heads cannot be analysed, hence they were discarded from the set of images. Luckily, due to the conditions under which the sample is obtained, overlapped heads do not appear frequently.

Each head was registered automatically in order to assure scale and rotation invariance. First of all, the heads were rotated to its vertical position. This was per-

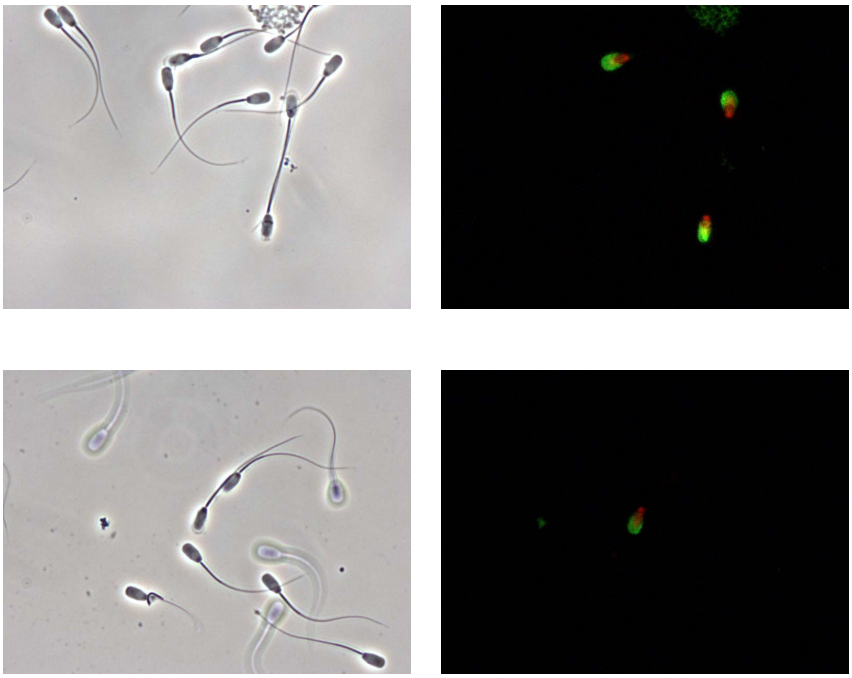


Figure 3.1: Sperm samples with intact and damaged acrosomes. (left) Phase contrast image. (right) Fluorescence image.

formed by fitting an ellipse to an sperm head and correcting the orientation of the major axis to achieve verticality. Then, the image is right and left cropped leaving head's pixels untouched. Afterwards, the coordinates of the tail were detected. The image was flipped when the tail was placed in the top half of the image. Then, the image was up and down cropped leaving head's pixels intact. Finally, all images were resized to the median dimensions of the set. Bad registered images were manually discarded.

Our dataset is composed by 856 intact and 861 damaged heads of 56×108 pixels. Figure 3.2 shows examples of intact and damaged registered sperm heads.

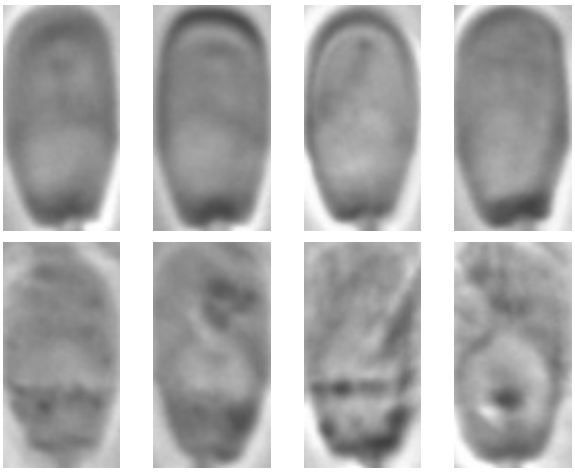


Figure 3.2: (top) Intact registered acrosomes. (bottom) Damaged registered acrosomes.

3.2. Invariant local features versus traditional texture descriptors

This section aims at comparing the performance of invariant local features (ILF) and traditional global texture descriptors when classifying boar spermatozoa heads as intact or damaged. SIFT and SURF are the ILF methods tested while Zernike moments, Haralick features extracted from the original image and from the coefficients of the discrete wavelet transform (DWT), Legendre moments and Laws masks make up the global texture descriptors analysed.

3.2.1. Method

Haralick

We extract 13 out of the 14 features proposed by Haralick (1979) from the original image, all except the maximal correlation coefficient following Alegre et al. (2009). These metrics are computed from the GLCM. This matrix is a second order texture statistic that represents how often different combinations of gray levels occur between two pixels of the neighbourhood at a given offset. Formally, if i and j are the image intensity values of the image, p and q are the spatial positions in the $n \times m$ image I and $(\Delta x, \Delta y)$ the offset dependent on the direction θ and the distance d at which the matrix is computed, the GLCM is obtained following the equation 3.1.

$$GLCM_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m}\begin{cases}1 & \text{if } I(p,q)=i \text{ and } I(p+\Delta x, q+\Delta y)=j \\ 0 & \text{otherwise}\end{cases} \tag{3.1}$$

We compute the GLCM in four directions $\theta = \{0^o, 45^o, 90^o, 135^o\}$ obtaining the average value in order to achieve rotation invariance. Moreover, we calculate the GLCM for distances $d = \{1, 2, 3, 5\}$.

WCF13

Information represented by spatial frequencies is often used for texture pattern recognition. We apply the DWT to the images. It localises the high-frequency components of a signal, in our case an image, so that they can be analysed separately. In order to obtain DWT from an image, two dimension DWT is considered, leading to a decomposition into four sub-bands in each level of decomposition, for the first level LL1, LH1, HL1 and HH1. These four sub-bands arise from applying horizontal and vertical (low (L) or high (H)) filters one after the other and down-sampling each filtering with a factor of 2. LL1 represents the course information of the image and it is used to obtain the next levels of wavelet coefficients, while the rest conform the detail images. We use Haar wavelets (Haar, 1910) in the computation of the DWT.

We compute the same 13 Haralick features on the GLCM of the original image and on the GLCMs of the four sub-images of the first level of decomposition with the Haar DWT, leading to a descriptor composed by 65 features which is called WCF13.

Laws masks

This method is based on texture energy transforms. Laws (1979) developed a set of two-dimensional masks derived from five simple one-dimensional vectors of five pixels length: level (L_5), edges (E_5), spots (S_5), ripples (R_5) and waves (W_5). By convolution of any vertical one-dimensional vector with a horizontal one two-dimensional masks of size 5×5 are generated. In our work we are only interested in level, edges, spots and ripples so we obtained 16 masks, Alegre et al. (2009). For example, the mask L_5E_5 is computed by convolving vertical L_5 and horizontal E_5 vectors.

First we normalize by subtracting from each pixel the average of its 15×15 neighbour in order to remove the effects of illumination. Then, we convolve the image $I(i,j)$ by a Laws mask X_5X_5, $J(i,j) = I(i,j) * X_5X_5$. We compute the energy maps

E with a moving non-linear window average of absolute values:

$$E(r,c) = \sum_{j=c-7}^{c+7} \sum_{i=r-7}^{r+7} |J(i,j)| \tag{3.2}$$

Finally, we combine the energy maps of certain symmetric pairs of filters, producing 9 descriptors which are: L_5E_5/E_5L_5, L_5S_5/S_5L_5, L_5R_5/R_5L_5, E_5E_5, E_5S_5/S_5E_5, E_5R_5/R_5E_5, S_5S_5, S_5R_5/R_5S_5 and R_5R_5.

Legendre moments

Moments are able to extract global features thus, they have been extensively applied in the field of image processing. Teague (1980) introduced Zernike moments and Legendre moments. This moments are orthogonal and as a result the reconstruction of an image from the mathematical features provided by these moments is possible. Shu et al. (2000) present an efficient method for computation of Legendre moments. We use 9 Legendre moments to describe the images: the ones from order (0, 0) to order (2, 2).

Zernike moments

Also Teague (1980) proposed Zernike moments based on the set of orthogonal Zernike polynomials. Particularly, Zernike moments have been shown to be rotation invariant and robust to noise. A relatively small set of Zernike moments can characterize the global shape of a pattern effectively. The low order moments represent the global shape of a pattern and the higher order the detail. When the Zernike moments of the image are suitable for a large number of terms, then the reconstruction of the input image function can be achieved with high accuracy. We obtained the absolute value of 9 Zernike moments (until 4th order).

SIFT

SIFT (Lowe, 2004) transforms image data into scale-invariant coordinates relative to local features. It can be described in 4 main stages:

- Scale-space extrema detection: The first stage of computation searches over all scales and image locations. It is implemented efficiently by using a difference-of-Gaussian function to identify potential keypoints that are invariant to scale and orientation.

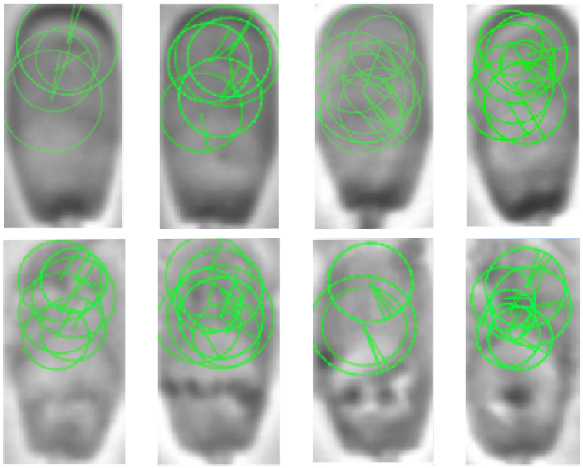


Figure 3.3: SIFT keypoints found in spermatozoa heads. Keypoints are displayed as green circles indicating scale, orientation and location at which they were found. (top) Intact registered acrosomes. (bottom) Damaged registered acrosomes.

- Keypoint localization: At each candidate location, a detailed model is fit to determine location and scale. Keypoints are selected based on measures of their stability.
- Orientation assignment: One or more orientations are assigned to each keypoint location based on local image gradient directions. All future operations are performed on image data that has been transformed relative to the assigned orientation, scale and location for each feature, thereby providing invariance to these transformations.
- Keypoint descriptor: The local image gradients are measured at the selected scale in the region around each keypoint. These are transformed into a representation that allows for significant levels of local shape distortion and change in illumination.

Feature descriptors have a 128 dimensionality for each keypoint of the image: 8 directions for each histogram orientation in 4×4 subregions. Figure 3.3 shows the keypoints localized with SIFT in both intact and damaged acrosomes.

SURF

SURF approach (Bay et al., 2008) for keypoint detection uses a very basic Hessian-matrix approximation ($\mathsf{H}(x, \sigma)$) defined in Eq.3.3 that relies on box filters

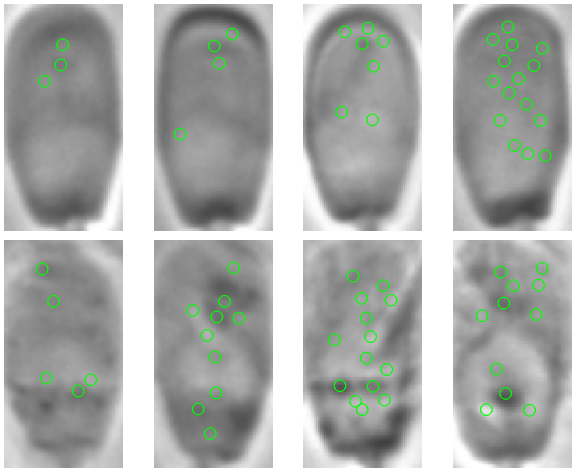


Figure 3.4: SURF keypoints found in spermatozoa heads. (top) Intact registered acrosomes. (bottom) Damaged registered acrosomes.

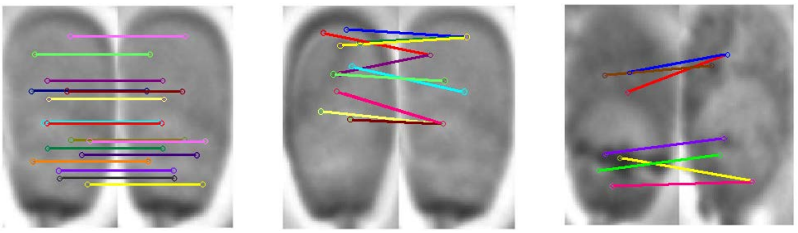
as approximations of the Gaussian second order derivatives. This lends itself to the use of integral images which reduces the computation time drastically. Keypoints are found at different scales.

$$\mathsf{H}(x,\sigma) = \begin{bmatrix} D_{xx}(x,\sigma) & D_{xy}(x,\sigma) \\ D_{xy}(x,\sigma) & D_{yy}(x,\sigma) \end{bmatrix} \tag{3.3}$$

For the extraction of the descriptor, the first step consists of constructing a square region centred around the keypoint and oriented along the reproducible orientation assigned. The region is split up regularly into smaller 4×4 square sub-regions and for each sub-region, Haar wavelet responses are computed. Then, the wavelet responses d_x and d_y in horizontal and vertical directions respectively are summed up over each sub-region and form a first set of entries in the feature vector. In order to bring in information about the polarity of the intensity changes, the sum of the absolute values of the responses, $|d_x|$ and $|d_y|$, is also extracted. Hence, each sub-region has a four-dimensional descriptor vector V for its underlying intensity structure:

$$V = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|) \tag{3.4}$$

Concatenating these descriptors for all 4×4 sub-regions, it results in a feature vector of length 64 for each key point of the image. In Fig 3.4 keypoints found with SURF in intact and damaged acrosomes are shown.



(a) (b) (c)

Figure 3.5: Matching correspondences of SURF keypoints between pairs of spermatozoa heads. (a) The same head. (b) Two intact acrosomes. (b) Two damaged acrosomes.

3.2.2. Experiments

We classify the test images using k-NN method by taking one element of the test set and finding its k nearest elements in the training set. The class assigned to that element is the most repeated one in those k elements. We used odd k values between 1 and 15 inclusive.

SIFT and SURF have as many descriptors per image as keypoints were found. We compute the distances from each descriptor in a test image to all descriptors in a training image and consider the match that achieves the minimum distance. We repeat this calculation for all descriptors in the test image and calculate the sum of the minimum distances as the distance between training and test images. Then, we apply k-NN.

The proximity between the patterns is computed using the Euclidean distance for all the evaluated methods. Let $X = (x_1, x_2, ...x_n)$ be a descriptor of the test set and $Y = (y_1, y_2, ...y_n)$ be a descriptor of the training set. The Euclidean distance between descriptors $d(X, Y)$ is:

$$d(X,Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^2 \right)^{1/2} \tag{3.5}$$

Figure 3.5 shows examples of the matching correspondences between spermatozoa heads using Euclidean distance and SURF descriptors. It is remarkable that, in intact acrosomes keypoints are mostly found on the top half of the head whereas in damaged acrosomes they are found on the whole image. This could be due to the irregular texture also present at the half bottom of damaged acrosomes.

Furthermore, we also compute the cosine similarity $cos(X, Y)$ for SIFT method since it is proposed by its authors (Lowe, 2004). It measures the cosine of the angle

between two vectors. The result of the cosine function is equal to 1 when the angle is 0, and it is less than 1 when the angle is of any other value. Calculating the cosine of the angle between two vectors thus determines whether these two vectors are pointing in roughly the same direction.

$$cos(X,Y) = \frac{X \cdot Y}{||X||||Y||} = \frac{\sum_{i=1}^{n} X_i \times Y_i}{\sqrt{\sum_{i=1}^{n} X_i^2}\sqrt{\sum_{i=1}^{n} Y_i^2}} \tag{3.6}$$

We randomly take a set of 70% of images of each class for training and the rest for testing. We consider a true positive (TP) when a damage acrosome is classified as damage, a false positive (FP) when an intact acrosome is classified as damaged, a false negative (FN) when a damaged acrosome is classified as intact and a true negative (TN) when an intact acrosome is classified as intact. We compute the accuracy of the classification as the rate of correct assignments over the whole test population accuracy $= (\text{TP} + \text{TN})/(\text{TP} + \text{FP} + \text{FN} + \text{TN})$. Since the dataset is quite balanced, accuracy is a suitable way to evaluate the performance of the method and it is useful to compare our results with the state of the art ones. This process is repeated 10 times in order to achieve robustness to random choices. The final accuracy is the average of the accuracy rates during those 10 runs. We also computed the accuracy of each class. We define accuracy of the intact class as $\text{accuracy}_{\text{intact}} = \text{TN}/(\text{TN} + \text{FN})$ and accuracy of the damaged class as $\text{accuracy}_{\text{damaged}} = \text{TP}/(\text{TP} + \text{FP})$.

3.2.3. Results

Figure 3.6 and Table 3.1 show the accuracy results of each evaluated method, both in global terms and for each class, for the number of neighbours k that achieved the maximum global accuracy. We present results only for the distance d of Haralick and WCF13 that reached the best global accuracy. Figure 3.7 shows the global accuracies for the different values of k obtained by each evaluated method. The highest accuracy was yielded by SURF with $k = 11$ (94.88%), outperforming the rest of descriptors for every value of k considered. It can also be noticed that SIFT results are nearly the same regardless of the metric used, Euclidean distance or cosine similarity. Nevertheless, Euclidean distance always performed slightly better.

SURF and SIFT obtained better accuracy rates for the damaged acrosomes than for the intact ones for any value of k evaluated as shown in Fig. 3.8 and 3.9. SURF yielded an accuracy equals to 92.89% on the intact class and to 96.86% on the damaged one. It is even more noticeable for SIFT, achieving accuracies of 76.15% and

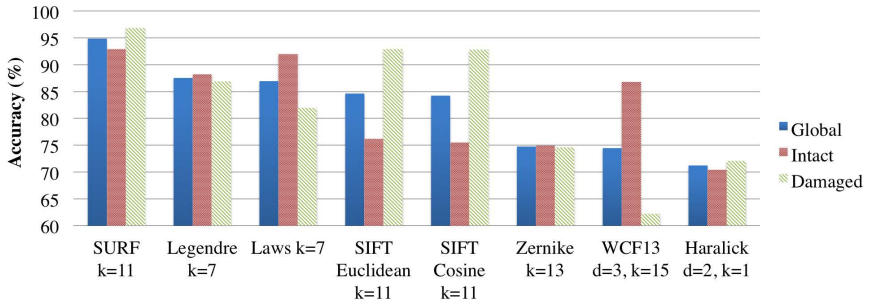


Figure 3.6: Best accuracies of each assessed method for the number of neighbours k that achieved the maximum global accuracy.

Table 3.1: Best accuracies of each assessed method for the number of neighbours k that achieved the maximum global accuracy.

Descriptors	k	Global (%)	Intact (%)	Damaged (%)
SURF	11	**94,88**	**92,89**	**96,86**
Legendre	7	87,55	88,24	86,86
Laws	7	86,95	91,98	81,94
SIFT Euclidea	11	84,64	76,15	92,96
SIFT cosine	11	84,24	75,52	92,80
WCF13 d=3	15	74,76	74,90	74,61
Zernike	11	74,46	86,80	62,21
Haralick d=2	1	71,22	70,39	72,05

92.96% for intact and damaged classes respectively using Euclidean distance. On the contrary, texture descriptors obtained better results on the classification of intact acrosomes with the exception of Haralick, Fig. 3.6 and Table 3.1. We believe that a combination of both global texture descriptors and ILF could improve their individual results.

3.3. SVM classification of SURF descriptors

3.3.1. Motivation

The fact that some ILF methods describe patches around the found keypoints in an image means that we have to deal with a variable number of descriptors per image depending on the number of keypoints detected. Therefore, many well-known

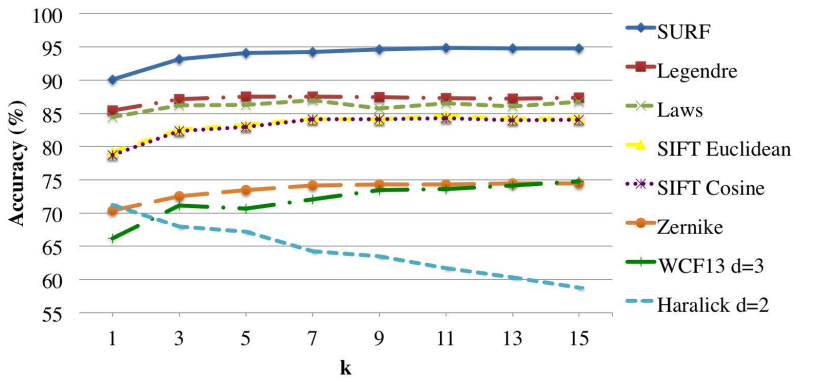


Figure 3.7: Results for the different values of neighbours k evaluated.

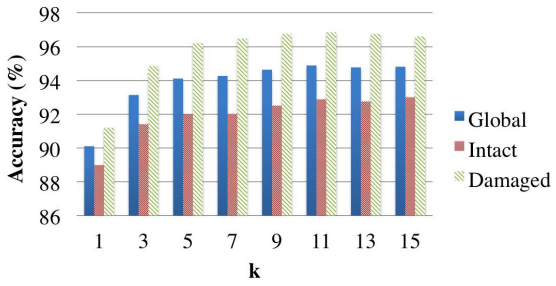


Figure 3.8: Results, both global and per class, using SURF with the different values of neighbours k evaluated.

classifying algorithms which work with one vector per image cannot be directly applied in this situation.

Usually, works dealing with ILF trust in nearest neighbours algorithms in order to classify keypoints descriptors (Lowe, 2004). This approach compares each test image with every training image and set the test image class as the most repeated class over the k most alike training correspondences. The correspondence between images does not rely on the correspondence of two feature vectors but of two feature matrices. One typical solution is to compute the minimum distance from one descriptor in a test image to all descriptors in a training image. Then a metric is applied to all minimum distances obtained for all descriptors in the test image. This metric can be the sum of the minimum distances, the absolute minimum distance or any other.

Nevertheless, no algorithms are known that can identify the exact nearest neigh-

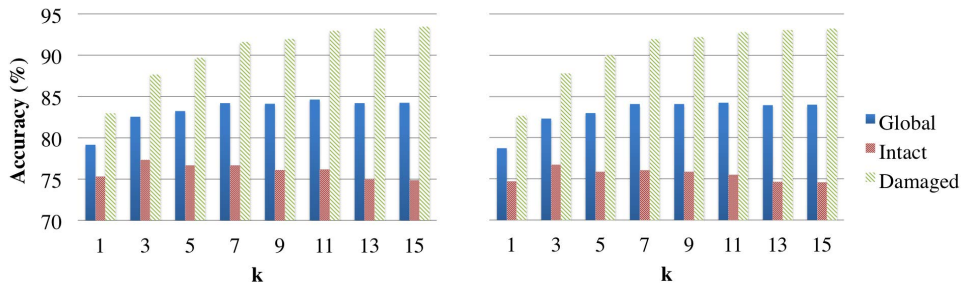


Figure 3.9: Results, both global and per class, using SIFT with the different values of neighbours k evaluated. (a) With Euclidean distance. (b) With cosine similarity.

bours of points in high dimensional spaces that are more efficient than exhaustive search. In order to overcome the disadvantages that the traditional k-NN classification technique presents, such as slow speed and low efficiency (Liu et al., 2008), we have chosen to adapt SVM when dealing with several features vectors per image. Some works build histograms (bags of words) from the ILF in order to achieve a fixed size vector for every image in the dataset (Sidibé et al., 2015; Favorskaya and Proskurin, 2015). However, we aim to use the descriptors without the need to cluster them.

3.3.2. Method

In this work, we use SVM to classify both individual SURF descriptors of spermatozoa heads and spermatozoa heads described using SURF as intact or damaged.

First, we concatenate all keypoints descriptors in a 17122×64 matrix, where each row represents a keypoint descriptor and each column represents a SURF feature. Additionally, we define a ground truth label vector of 17122 elements in which each point belonging to an intact head is labelled as intact and vice versa. Finally, we define a vector of 17122 elements in which we give a unique identification number to each descriptor according to the spermatozoa head to which they belong. Figure 3.10 represents such a matrix and its labelling.

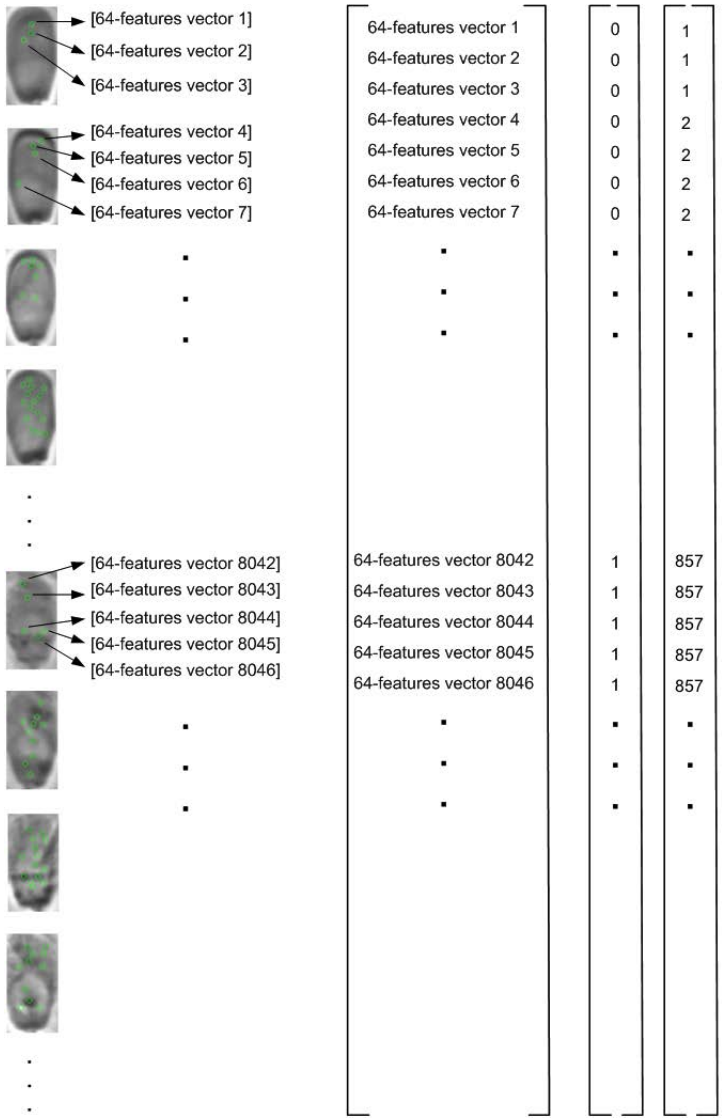


Figure 3.10: Each point detected in a head is described by a 64 feature vector using SURF (left). All points descriptors are orderly stored in rows into a matrix (middle). For each point, we create a vector giving the same labels to the keypoints as the labels of their spermatozoa heads and another vector with the identification numbers of the heads in which they were found (right).

3.3.3. Experiments

SVM applied to keypoints

Spermatozoa heads with damaged acrosome present regions such as black dots that could be easily located as corners, Fig. 3.2). This hypothesis implies that damaged class visually present potential keypoints different from possible keypoints in the intact class. First, we consider the classification of individual SURF descriptors of the spermatozoa heads. We use the ground truth created for each descriptor and the descriptors themselves of a training set in order to train a SVM with a linear least squares algorithm, Fig. 3.11a. We carry out a k-folds cross-validation with $k = 10$ for all keypoints. We compute the accuracy as the rate of correctly classified keypoints on the whole test set. Finally, we average the accuracy results for the 10 splits.

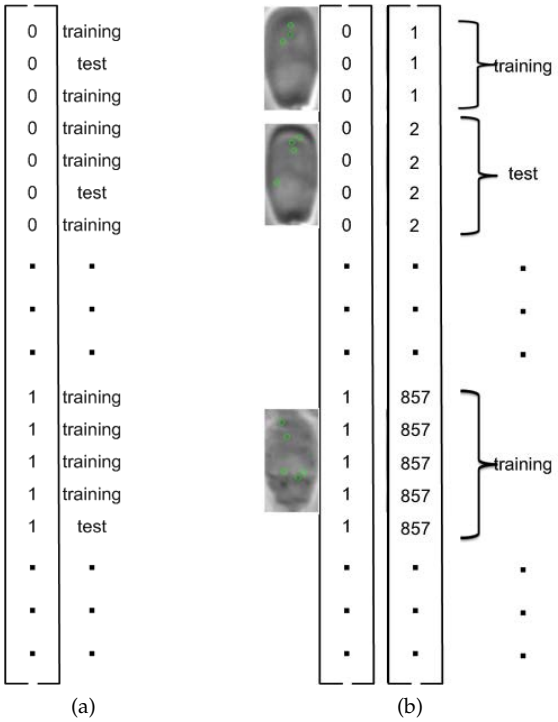


Figure 3.11: (a) k-folds applied to keypoints. (b) k-folds applied to heads.

SVM applied to heads

Spermatozoa heads present distinctive enough keypoints which can be undoubtedly seen either as intact or as damaged points whereas there are some others found in both classes that could be mistaken, Fig. 3.2. Nevertheless, it seems that, in general, an intact head contains more distinctive intact points than doubtful points and the analogous situation happens with damaged heads.

We implement again a SVM using a linear least squares training for individual descriptors. However, this time we perform the 10-fold cross-validation on the heads rather than on the keypoints, Fig. 3.11b. Consequently, descriptors belonging to 90% of the total amount of heads, no matter how many keypoints were found in each one, are selected to train our classifier in each run. Besides, now we measure the performance in terms of accuracy in the classification of heads rather than of keypoints. We consider a head as correctly classified when it yields a greater number of keypoints well classified by the SVM than the amount of mismatched keypoints. We compute the accuracy as the rate of correctly classified heads on the whole test set, see Section 3.2.2. We finally average the accuracy results for the 10 runs.

3.3.4. Results

Figure 3.12 shows the accuracies obtained with the two proposed approaches. Accuracies for each class -intact and damaged acrosomes- are also plotted. We just obtained an accuracy of 72.57% when classifying individual keypoints. When classifying heads, an overall accuracy of 90.91% was achieved, which represents an improvement of 25.27% over the former approach.

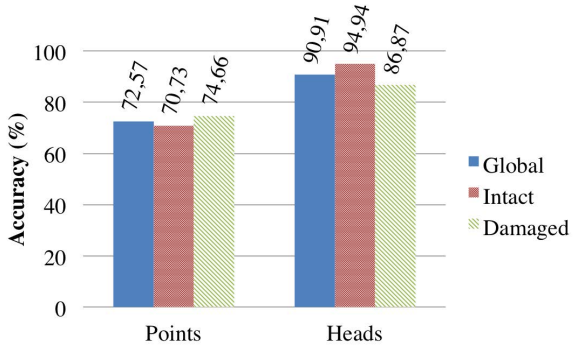


Figure 3.12: Global, intact and damaged accuracies using SURF and SVM applied to keypoints and to heads.

It is also remarkable that when classifying points, damaged heads were better classified than intact ones, whereas when heads are considered, the opposite situation was yielded. One reason could be that while damaged points are in general more distinctive than intact points, damaged heads contain areas where it is clear that the acrosome is damaged together with some areas where it is not appreciable and hence those latter areas could be considered as intact and lead to misclassifications.

This approach can be widespread to different invariant local features descriptors (SIFT, BRISK, FREAK, etcetera) and to other conventional classifying algorithms such as neural networks.

3.4. Combining ILF and global texture descriptors

In this section we aim at evaluating the performance of the combination of one local and one global description methods for the classification of the state of boar acrosomes as intact or damaged.

3.4.1. Method

We use SIFT and SURF methods as ILF descriptors and Legendre and Zernike moments, Laws masks and Haralick features as global texture descriptors already explained in Section 3.2.1. Therefore for each image, we perform an early fusion of a set of ILF descriptors with one global texture descriptor. We choose to concatenate all ILF descriptors of an image with the global texture descriptor of the same image. Thus, the matching process is directly affected by the dimensionality of the original descriptors. Before fusion, we normalize the individual descriptors to mean 0 and standard deviation 1.

3.4.2. Experiments

We use k-NN and SVM algorithms to carry out the classifications. k-NN is implemented as explained in Section 3.2.2.

SVM is applied by means of a bag of words (BoW) model (Aldavert et al., 2010; Li et al., 2011). BoW represents an image by the histogram of local patches on the basis of a visual vocabulary. First, it obtains n centres using a clustering algorithm taking into account all the image descriptors in the training set. We apply BOW with k-means clustering algorithm for $k = \{2, 3, ..., 10\}$. Secondly, the proximity between descriptors and centres is computed and each descriptor is assigned to its closest centre. We use Eculidean distance. The output of a BoW model for an image is a

histogram (vector) of size n where each element n_i represents the number of image descriptors assigned to the centre i. We use the BoW histograms of the training set images to train a SVM with a linear least squares algorithm. Finally, we compute the accuracy obtained in the classification of the test set. We define accuracy as in Section 3.2.2.

As for the k-NN, we randomly take 70% of the images of each class for training and the rest for testing. We repeat this process 10 times and the final accuracy rates are the average of the accuracies for those 10 runs.

3.4.3. Results

Figure 3.13 shows the results obtained with the early fusion of ILF and global texture descriptors classifying with k-NN and SVM using BoW. Best results were obtained when combining ILF descriptors with Legendre and classifying with k-NN. The best overall accuracy of 95.56% was reached with the combination of SURF and Legendre. Fusioning SIFT with Legendre yields an accuracy of 88.98%. These combinations improved the results yielded with each individual method using the same classification method (94.88%, 84.86% and 87.55% for SURF, SIFT and Legendre respectively) as presented in Section 3.2.3.

SVM with BoW obtained lower accuracies than k-NN. Combining SURF and Laws reached an accuracy of 75.66% whereas fusioning SIFT and Laws achieved 75.97%, in both cases with 4 centres in the dictionary. The low number of keypoints detected for these low resolution images (56×108 pixels) may cause a poor definition of the dictionary. Therefore, using BoW for this application is not discriminatory enough.

3.5. Conclusions

The contributions of the work presented in this chapter are four-fold. First, we demonstrated the success of applying a computer vision approach based on ILF for the evaluation of the state of boar acrosomes as intact or damaged. Second, we compared the performance of SURF and SIFT against some global texture descriptors for the application at hand. Third, we proposed a method to classify SURF features, which produces several descriptors per image, with traditional SVM classifiers. This approach can be easily implemented for other ILF and classifiers. Forth, a novel early fusion of ILF with global texture descriptors was presented improving previous works in this area of knowledge.

To our knowledge and to the date of publication of this work in different conferences, these were the best results achieved in the literature for the classification

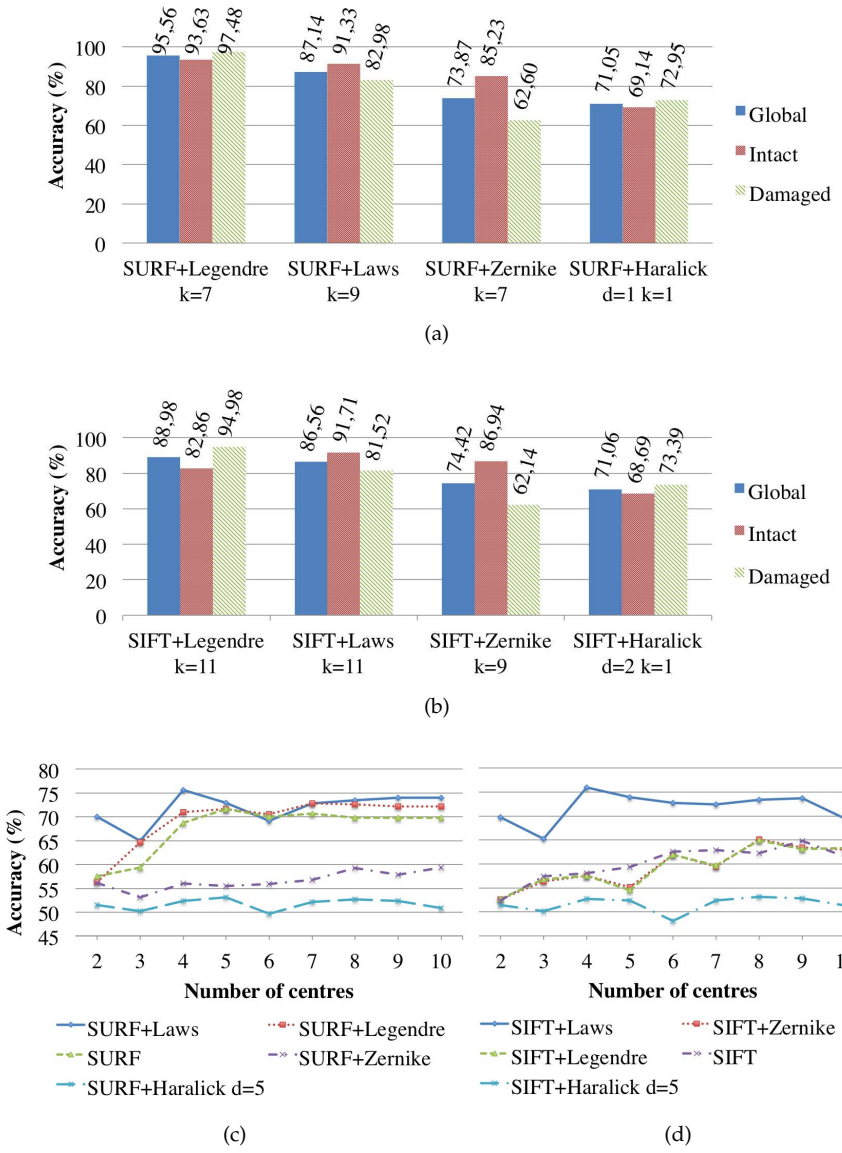


Figure 3.13: Results of the early fusion of ILF with global texture descriptors. (a) SURF as ILF and classifying using k-NN. (b) SIFT as ILF and classifying using k-NN. (c) SURF as ILF and classifying using BoW and SVM. (d) SIFT as ILF and classifying using BoW and SVM.

of boar spermatozoa as intact or damaged in such a big and balanced dataset. The

best result (accuracy equals to 95.56%) is satisfactory enough for the purpose of the application according to the veterinarian community.

Chapter 4

Automatic localisation of broken inserts in edge profile milling heads

4.1. dataset

To the best of our knowledge, there are no publicly available image datasets of milling cutting heads in the literature. For this reason, we created a new dataset with ground truth and we published it on-line[1]. It is made up of 144 images of an edge profile cutting head used in a computer numerical control (CNC) milling machine. We set up a capturing system as shown in Fig. 4.1. The head tool, with cylindrical shape, contains 30 inserts in total from which 7 to 10 inserts are seen in each image of the dataset. The inserts are arranged in 6 groups of 5 inserts diagonally positioned along the axial direction of the tool perimeter, as seen in Fig. 4.2. The last insert of each group is vertically aligned with the first insert of the following group. It gives a total of 24 different positions along the radial perimeter of the tool head in which at least one insert is aligned with the camera in intervals of 15°. Therefore, the same insert is captured in several images (between 7 and 9) under different poses as the head tool rotates, Fig. 4.3. The evaluation of inserts is planned to be performed during the resting state of the milling head tool between the processing of two metallic plates. The described capturing system can be set up at that resting position.

We created the dataset following an iterative process. We mounted 30 inserts in the head tool and took 24 images of the head tool in different orientations that differ by 15°. We repeat this process for 6 times, where each time we use a different set of inserts, thus collecting ($6 \times 24 =$) 144 images that contain ($6 \times 30 =$) 180 unique inserts, of which 19 are broken and 161 are unbroken. All inserts that we used to create this dataset were taken after some milling operations by the same machine.

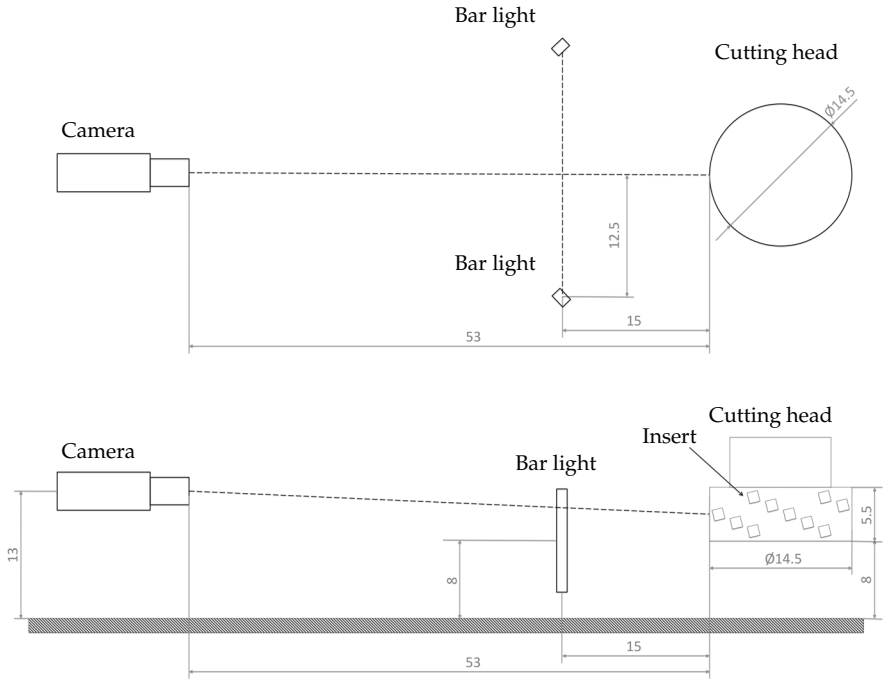
We used a monochrome camera Genie M1280 $1/3''$ with pixel size of 3.75 μm, active resolution of 1280×960 pixels and fixed C-mount lent AZURE-2514MM with a focal length of 25 mm and $2/3''$ format. The two compact bar shape structures

[1] http://www.computervisiononline.com/dataset/edge-milling-heads

with high red intensity LED arrays BDBL-R(IR)82/16H were used to enhance the image capturing capability and intensified the lighting on the edges. The milling machine that we used to create our dataset does not use oils, lubricants or other kind of substances that can cause a filthy tool.

Together with the dataset, we provide the corresponding ground truth masks of all ideal cutting edges along with the labels of the state of the inserts (broken or unbroken). Moreover, we labelled each distinct insert, by giving them unique identification numbers. In Fig. 4.3, we show three consecutive images that contain the same inserts (with the same identification number) in different locations and poses due to the rotation of the milling head in steps of $15°$.

Furthermore, we create a ground truth localising the centres of all inserts. Only complete inserts have been considered for creating this ground truth, discarding partly visible inserts. The ground truth consists of two parts. First, a list of coordinates for each central point of the screws that fasten the inserts. Secondly, for each



(a)

Figure 4.1: Front (top) and side (bottom) view of the capturing system. Measurements are in centimetres.

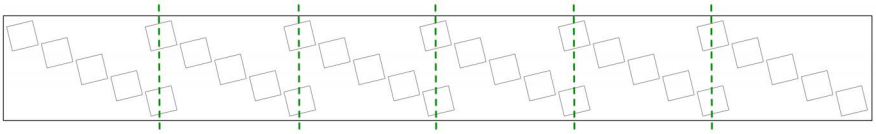


Figure 4.2: Diagram representing a schema of the arrangement of inserts on a cylindrical milling head depicted opened up as a rectangle. Squares denote the inserts. The vertical dashed line shows the alignment between adjacent groups of inserts.

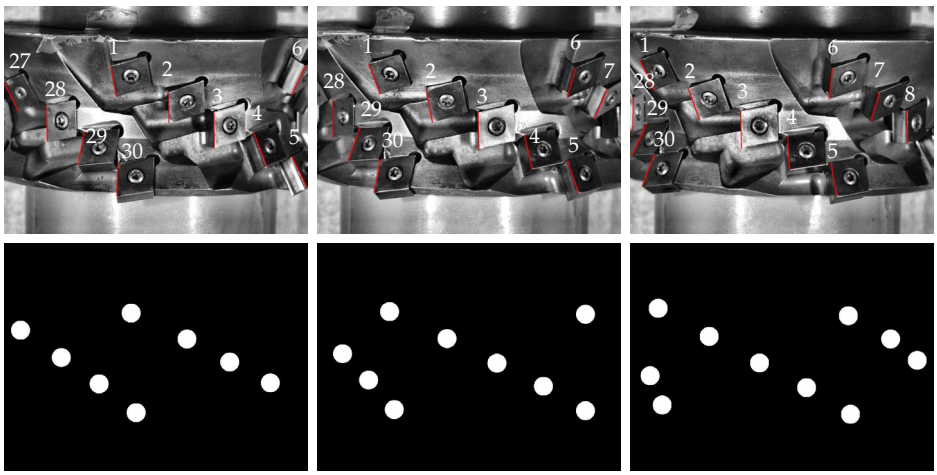


Figure 4.3: In the first row, the numbers indicate the ground truth labels of each cutting edge along three consecutive images of the dataset. Consecutive images are taken by rotating the milling head by 15°. A cutting edge present in different images is labelled with the same number. In the second row, ground truth circle masks located at the centres of the screws. The white circles approximately cover the screws of the image.

image of the dataset we provide a mask image with circles of radii 40 pixels centred at the previous coordinates, Fig. 4.3 second row. A radius of 40 pixels covers about a whole screw for all the images.

4.2. Automatic localisation of inserts and cutting edges using image processing

In this section we propose a methodology for the automatic detection of a region of interest (ROI) around the cutting edges of inserts that can be used to evaluate

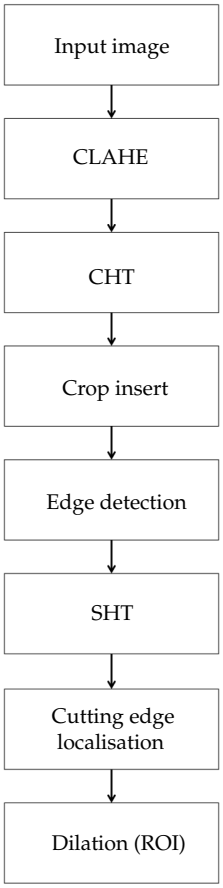


Figure 4.4: Outline of the proposed methodology.

their wear state at a later stage.

4.2.1. Method

The localisation that we propose is done in two steps. First, we detect the screws of the inserts and use them as reference points, and then we localise the cutting edges. In order to improve the quality of the images and facilitate the detection of edges, we apply the contrast-limited adaptive histogram equalization (CLAHE) method Zuiderveld (1994). Figure 4.4 shows a schema with all the steps in the proposed methodology. Below we elaborate each one of them.

Detection of inserts

The screw that fastens each insert has a distinctive circular shape. We use a circular Hough transform (CHT) to detect circles with radii between 20 and 40 pixels, because this is the size in which a screw appears on the images of size 1280×960 pixels. For the CHT, we use a two-stage algorithm to compute the accumulator array Atherton and Kerbyson (1999) Yuen et al. (1989). In the bottom row of Fig. 4.5 we show the CHT accumulator arrays for the images in the top row. By means of experiments, we set the sensitivity parameter of the CHT accumulator array to 0.85. The range of the sensitivity parameter is $[0, 1]$, as you increase the sensitivity factor, more circular objects are detected. Figure 4.5 shows examples in which the detected circles are marked in blue. Screws that appear in the left and right peripheries of the image are usually missed due to their elliptical shape. This does not pose a problem because the same insert is seen in different positions in the previous or next images.

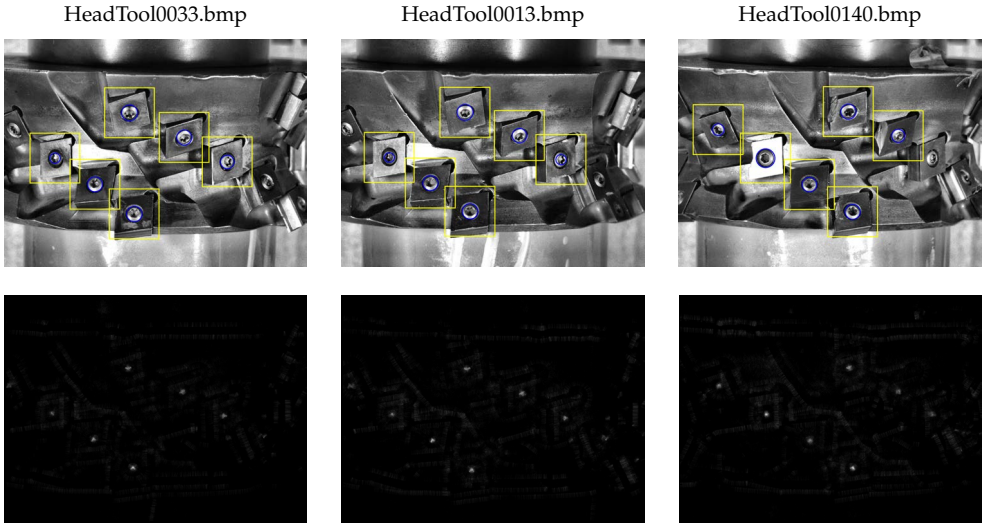


Figure 4.5: Firs row: In blue, detected circles by CHT. The circles are drawn with the detected radii and positioned around coordinates that have local maximum values. In yellow, cropped areas around the centre of the detected circles that contain a whole insert. Second row: Accumulator arrays obtained with CHT on the three images in top row.

We crop a rectangular area of size 205×205 pixels centred on a detected screw, the chosen dimensions are just enough to contain the whole insert. We then use this cropped area to identify the cutting edge. Figure 4.5 shows examples of cropped areas marked with yellow squares.

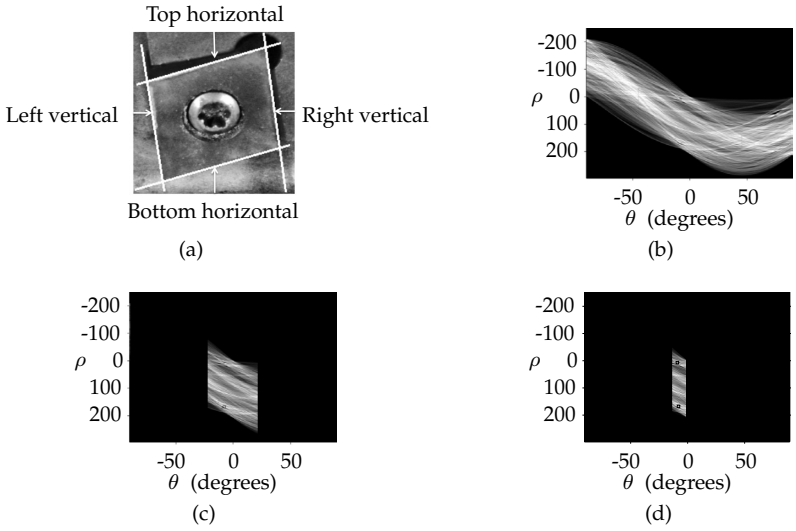


Figure 4.6: (a) Automatically detected lines that form the rhomboid shape of an insert. (b) Hough transform of the image in (a). (c) Hough transform for nearly vertical lines ($\pm$ 22^o). The black rectangle indicates the position of the largest peak ($\theta = -8^\circ$ and $\rho = 168$). (d) Hough transform for vertical lines with slope $(-8\pm 5)^o$. Black rectangles superimposed to the hough transform indicate two peaks that are greater than a fraction 0.75 of the maximum: ($\rho_1 = 7, \theta_1 = -9^\circ$) and ($\rho_2 = 168, \theta_2 = -8^\circ$).

Localisation of cutting edges

Inserts have a rhomboid shape formed by two nearly vertical ($\pm$ 22^o) and two nearly horizontal ($\pm$ 20^o) line segments (Fig. 4.6a).

First we use Canny's method (Canny, 1986) to detect edges in a cropped area (Fig. 4.7(a-b)). Then, we apply a standard Hough transform (SHT) (Hough, 1962) to the edge image in order to detect lines (Fig. 4.6(b-d)).

We look for the strongest vertical line segment which is represented as the highest value of peaks in the Hough transform matrix. Then, we look for line segments with peak values greater than a fraction 0.75 of the maximum peak value and with slopes in a range of $\pm$ 5^o with respect to the slope of the strongest nearly vertical line. In Fig. 4.6b we show the Hough transform of the cropped area shown in Fig. 4.6a. We consider the strongest nearly vertical line segment which is at least 47 pixels to the left of the center as the left edge of the insert. This detected line segment is considered as a full line and it is drawn in magenta in Fig. 4.7d. In this way, we avoid possible detection of lines around the screw area.

Similarly, we look for two horizontal line segments above and below the screw.

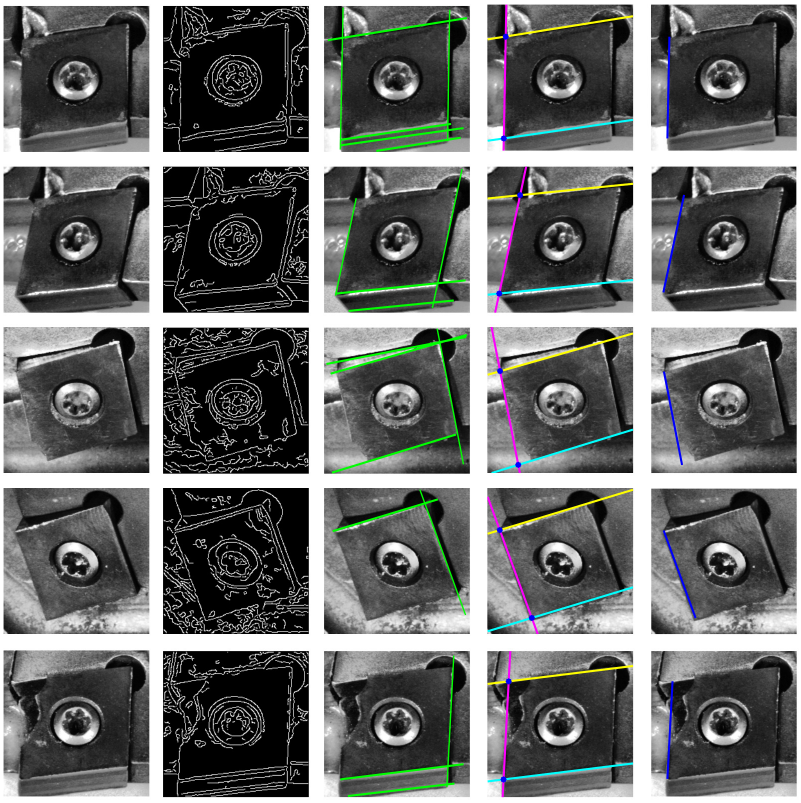


Figure 4.7: (a) Cropped areas containing inserts. (b) Canny edge maps. (c) Detection of (nearly) vertical and (nearly) horizontal lines. (d) Blue spots indicate the intersections between the two horizontal lines and the left vertical line. Lines obtained by symmetry are the following. Second row: top horizontal line; third row: left vertical line; forth row: left vertical line and bottom horizontal line; fifth row: left vertical line and top horizontal line. (e) Detected cutting edges.

In this case, the minimum distance from the line to the centre is set to 52 pixels and the range of possible slopes is $\pm\ 11^o$ with respect to the slope of the strongest horizontal line. The top and bottom detected lines are shown in yellow and cyan respectively in Fig. 4.7d. The points where the horizontal lines intersect with the left vertical line define the two ends of the cutting edge segment. These points are marked as dark blue dots in Fig. 4.7d. The localised cutting edges in these examples are shown in Fig. 4.7e.

The fact that we have a controlled environment (fixed camera and fixed resting

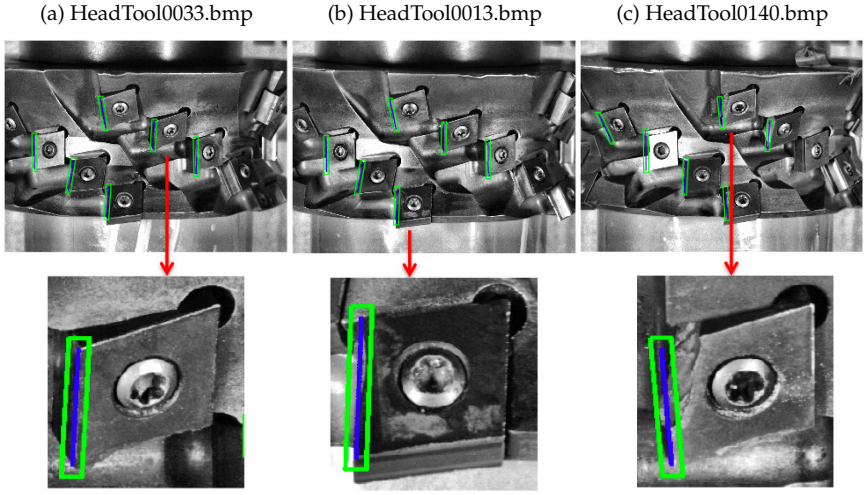


Figure 4.8: The blue segments define the localized cutting edges. The green frames mark the ROI that is achieved by means of a morphological dilation. (a) Image in which cutting edges are intact. (b) Image with some worn cutting edges. (c) Image with some broken inserts.

position of the head tool), this set of parameters is fixed only once for every given milling machine.

If the left line segment or any of the horizontal segments are not detected, we use symmetry to determine the missing lines. For instance, if the vertical line on the left of the screw is not detected but the one on the right is detected, we reconstruct the left line by rotating by 180 degrees the left line around the center of the concerned area. The bottom three examples in Fig. 4.7 show this situation.

Finally, we define a ROI by dilating the detected cutting edge segment with a square structuring element of 10 pixels radius. In Fig. 4.8, we show the cutting edge segments and ROIs localised by the proposed method for images containing inserts with different wear state. Notably is the fact that the proposed method can generalise the localisation of the cutting edge even in cases of worn or broken inserts.

4.2.2. Experiments and results

For each of the input images, we determine a set of ROIs around the identified cutting edges using the method described in Section 4.2.1. If the ground truth of a cutting edge lies completely in a ROI, we count that ROI as a hit and when it does not lie within any of the determined ROIs, the hit score is 0. If the ground truth overlaps a ROI, the hit score is equal to the fraction of the ground truth segment

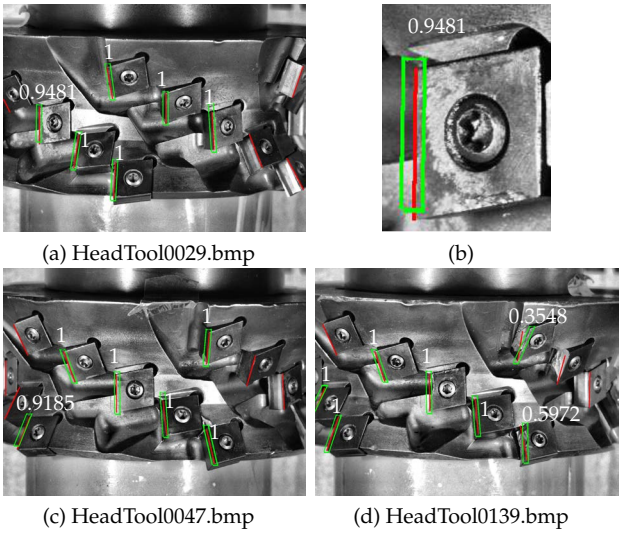


Figure 4.9: The green quadrilaterals are the ROIs detected by the proposed method and the red lines represent the ground truth for the cutting edges. The accuracy scores of the inserts are indicated in white font. (b) Example of a cutting edge that is not completely contained within the detected ROI. The accuracy of 0.9481 is the fraction of pixels of the cutting edge that lie within the detected ROI.

that lies inside a ROI. Some examples can be observed in Fig. 4.9.

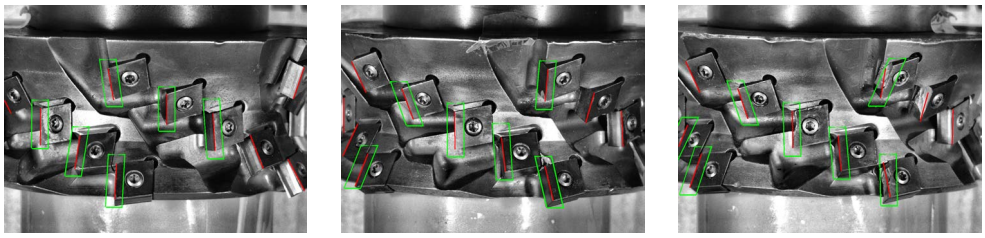
Every insert is detected in at least one of the 144 images. Moreover, whenever an insert is detected, the corresponding cutting edge on the left side is also always determined. We measure the accuracy of the method as an average of the partial scores for the individual cutting edges. Using this protocol, we obtain an accuracy measure of 99.61%.

Results can be improved by increasing the width of the structuring element in the final dilation stage. With a square structuring element of radius 34 pixels we achieve 100% accuracy. Figure 4.10 shows examples of the resulting ROIs.

4.2.3. Discussion

To the best of our knowledge the proposed approach is the first one automatically localises multiple inserts and cutting edges in an edge profile milling head.

Parameters have been computed in order that they can generalise for every insert at any position in the milling head tool given the geometry of the head tool and the arrangement of the capturing system. For a specific milling machine, parameters



HeadTool0029.bmp HeadTool0047.bmp HeadTool0139.bmp

Figure 4.10: Red line segments define the ground truth and green quadrilaterals define the detected ROIs with a morphological dilation operation of 34 pixel radius square structuring element.

can be easily estimated and then no further need of adjustment is needed.

We achieve an accuracy of 99.61% for the detection of cutting edges. This is achieved by dilating the automatically detected line using a square structuring elements of 20 pixel side. When the quadrilateral is 68 pixels wide, the accuracy reaches 100%. In future works, the ROIs defined around the detected cutting edges can be used for further evaluation of the wear state of the cutting edges.

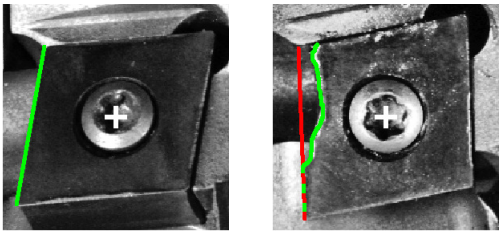
Furthermore, the proposed method can be used for different milling heads containing polygonal inserts fastened by screws, a design which is typical in edge milling machines. We implemented the proposed approach in Matlab and ran all experiments on a personal computer with a 2 GHz processor and 8 GB RAM. It takes less than 1.5 seconds to process all the steps on one image it takes about 1 minute to capture and process the 24 images taken to the head tool. This milling head tools are resting between 5 to 30 minutes, so the implementation reaches real time performing.

4.3. Classification of inserts as broken or unbroken

In this section we present a method for the classification of inserts as broken or unbroken by analysing the cutting edges that have been already localised in Section 4.2.1.

4.3.1. Method

In the method that we propose we first localise cutting edges in a given image, and then we classify every cutting edge as *broken* or *unbroken*. From the image



(a) (b)

Figure 4.11: (a) In green the real cutting edge of an intact insert. The white cross marks the centre of the detected screw. (b) In green the real cutting edge of a broken insert. In red the ideal cutting edge. All markers are manually drawn.

analysis point of view, an unbroken insert is one which has a straight cutting edge (Fig. 4.11a), while a broken insert has a curved or uneven cutting edge (Fig. 4.11b).

Figure 4.12 presents a schema that shows the proposed methodology. First we localise the inserts and the cutting edges and then, we evaluate the inserts using a three-stage method: applying an edge preserving smoothing filter, computing the gradient for each edge and finally using geometrical properties of the edge to assess its state. Below we elaborate on each of these steps.

Detection of inserts and localisation of ideal cutting edges

We use the algorithm introduced in Section 4.2 to detect inserts and localise the respective ideal cutting edges. For each localised insert, we consider a set I of Cartesian coordinates that form the ideal cutting edge.

$$I = \{(x_t, y_t) \,|\, t = 1...u\} \tag{4.1}$$

where u is the number of locations of the ideal cutting edge of a localised insert.

We determine a region of interest (ROI) from the ideal cutting edge and the horizontal edges that are detected by the algorithm in Section 4.2. In Fig. 4.13a we show examples of ROIs in broken and unbroken inserts. A ROI is determined by considering two parallel lines to the ideal cutting edge, one 3 pixels to the left and the other one to the right with a distance of 0.7 times the space between the ideal cutting edge and the centre of the screw. Moreover, we consider a parallel line to the top edge 3 pixels towards the bottom and a parallel line to the bottom edge 3 pixels towards the top. From the resulting quadrilateral, we remove a segment from a circle (with a radius of 45 pixels) around the centre of the screw that coincides with the quadrilateral. Such a ROI is sufficient to evaluate the state of a cutting edge while ignoring

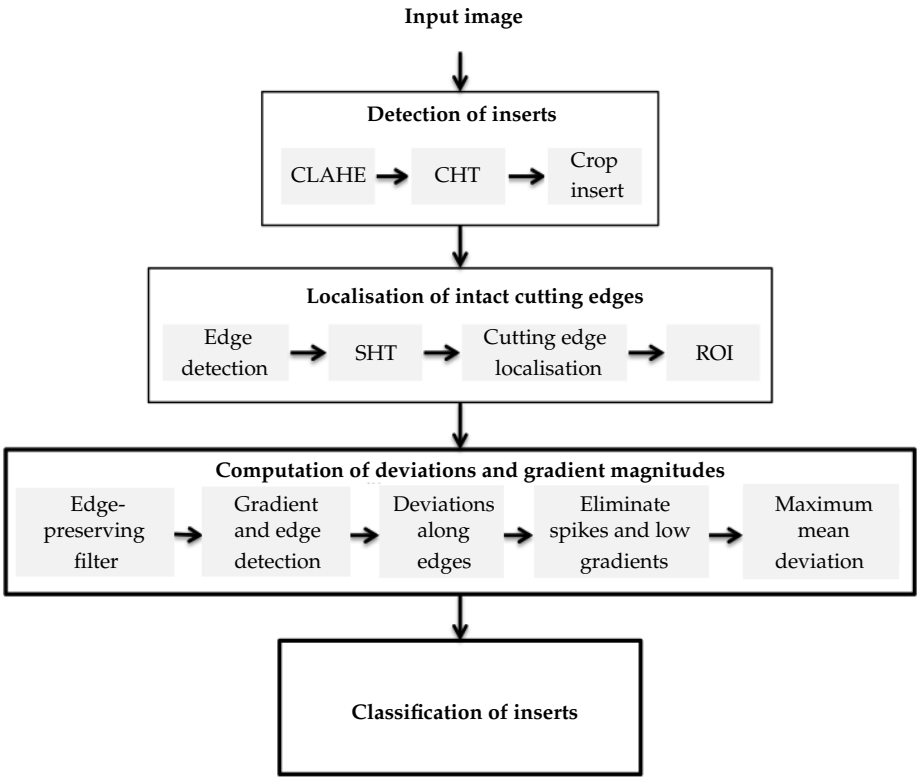


Figure 4.12: Outline of the proposed methodology.

possibly worn edges coming from the top or bottom parts of the insert as well as ignoring any texture coming from the screw. In the end, we consider a rectangle around the ROI with a 3-pixel width boundary and use it to crop the corresponding part of the image that contains the ROI, Fig. 4.13b. We also consider a mask defining the ROI in such a rectangular area, Fig. 4.13c.

Detection of real cutting edges

The heterogeneous texture and the low contrast of the insert with respect to the head tool make the detection of the real cutting edge an arduous task. If an edge detector is applied directly to the cropped images, many edges apart from the cutting edge would be recognised. In order to enhance the edge contrast, we apply

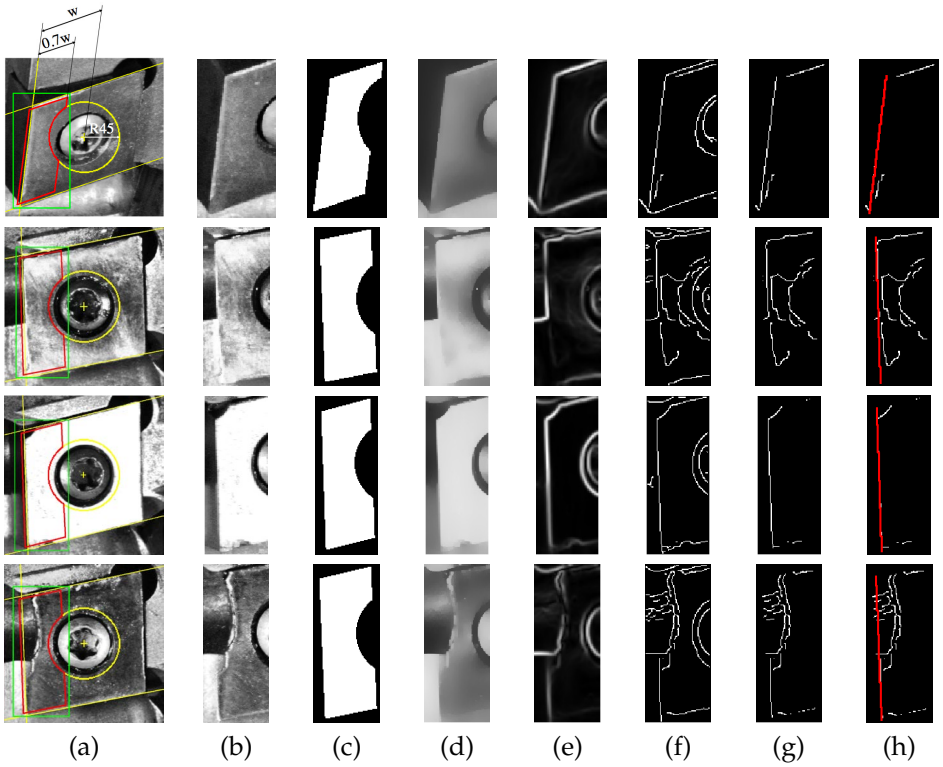


Figure 4.13: (a) In yellow, ideal, top and bottom edges. In red, definition of the ROI. In green, rectangle to crop. (b) Cropped region. (c) Mask defining the ROI in a cropped region. (d) Edge-preserving smoothed region. (e) Gradient magnitude map. (f) Edge map. (g) Result of multiplying the edge map by the mask. R set. (h) R set in white and I set overlaid in red. The two top inserts are unbroken while the two bottom ones are broken.

the edge-preserving smoothing filter of Gastal and Oliveira (2011)[2] to the cropped region. We choose this approach due to its efficiency and its good performance. This filtering method smooths the heterogeneous texture of the insert and of the background but preserves the edges of the insert. The images in Fig. 4.13d show examples of the output of this algorithm.

Afterwards, we apply Canny's method Canny (1986) to find edges by looking for local maxima of the gradient on the filtered region. Other edge detectors, such as the ones based on Sobel, Prewitt, Roberts and LoG, performed worse. Canny's algorithm computes the gradient after applying a Gaussian filter that reduces noise.

[2]Standard deviation of the spatial filter equals 60 and standard deviation of the range filter equals 0.4, as the default configuration.

Non-maximal suppression is applied to thin the edge. This is followed by hysteresis thresholding which uses a low and a high threshold in order to keep the strong edges (above the high threshold) and only the weak edges (with a value between the low and high threshold) that are connected to any of the strong ones. We show examples of Canny's gradient magnitude and binary edge maps in Fig. 4.13(e-f). Finally, we only consider the edges within the ROI (Fig. 4.13g). For each localised insert, we define a set R of $3-$tuples that represent the Cartesian coordinates (x_q, y_q) and the corresponding gradient magnitude value g_q of each location in the real cutting edge:

$$R = \{(x_q, y_q, g_q) \,|\, q = 1...v\} \tag{4.2}$$

where v is the number of locations of the real cutting edge of the localised insert.

Measurement of deviations between real and ideal cutting edges

For a pair of coordinates (x_t, y_t) in the ideal set of edges I, we determine a set P_t of coordinates $(\hat{x}, \hat{y})$ and the corresponding gradient magnitudes $\hat{g}$ from the set of real edges R such that $(\hat{x}, \hat{y})$ lie on a line that passes through (x_t, y_t). The slope m of this line is the gradient of the top edge.

$$P_t = \{(\hat{x}, \hat{y}, \hat{g}) \mid \hat{y} = m(\hat{x} - x_t) + y_t,\ (\hat{x}, \hat{y}, \hat{g}) \in R,\ (x_t, y_t) \in I\} \tag{4.3}$$

Examples of such lines are marked in blue in Fig. 4.14a. Next, we denote by E_t the set of Euclidean distances from (x_t, y_t) to all coordinates in the set P_t:

$$E_t = \left\{ \sqrt{(x_t - \hat{x}_p)^2 + (y_t - \hat{y}_p)^2} \;\middle|\; (x_t, y_t) \in I,\ \forall\, (\hat{x}_p, \hat{y}_p) \in P_t \right\} \tag{4.4}$$

E_t could be an empty set. Let D be the set of minimum distances of E_t for each point t in I. D represents the minimum deviations between the ideal and real edges.

$$D = \{\min(E_t) \,|\, t = 1...|I|\} \tag{4.5}$$

Let G be the set of gradient magnitudes of the points in P_t with the minimum distance in E_t for each point in I.

$$G = \{g_i \mid\ g_i \in P_t,\ i = \operatorname{argmin}(E_t),\ t = 1...|I|\} \tag{4.6}$$

In Fig. 4.14 we plot the values of the sets D and G.

We remove abnormal deviations that are usually caused by texture on the surface of the insert rather than by the cutting edge. For example, Fig. 4.14(e) presents two such abnormal deviations (spikes) at the beginning and end of the set D. We denote

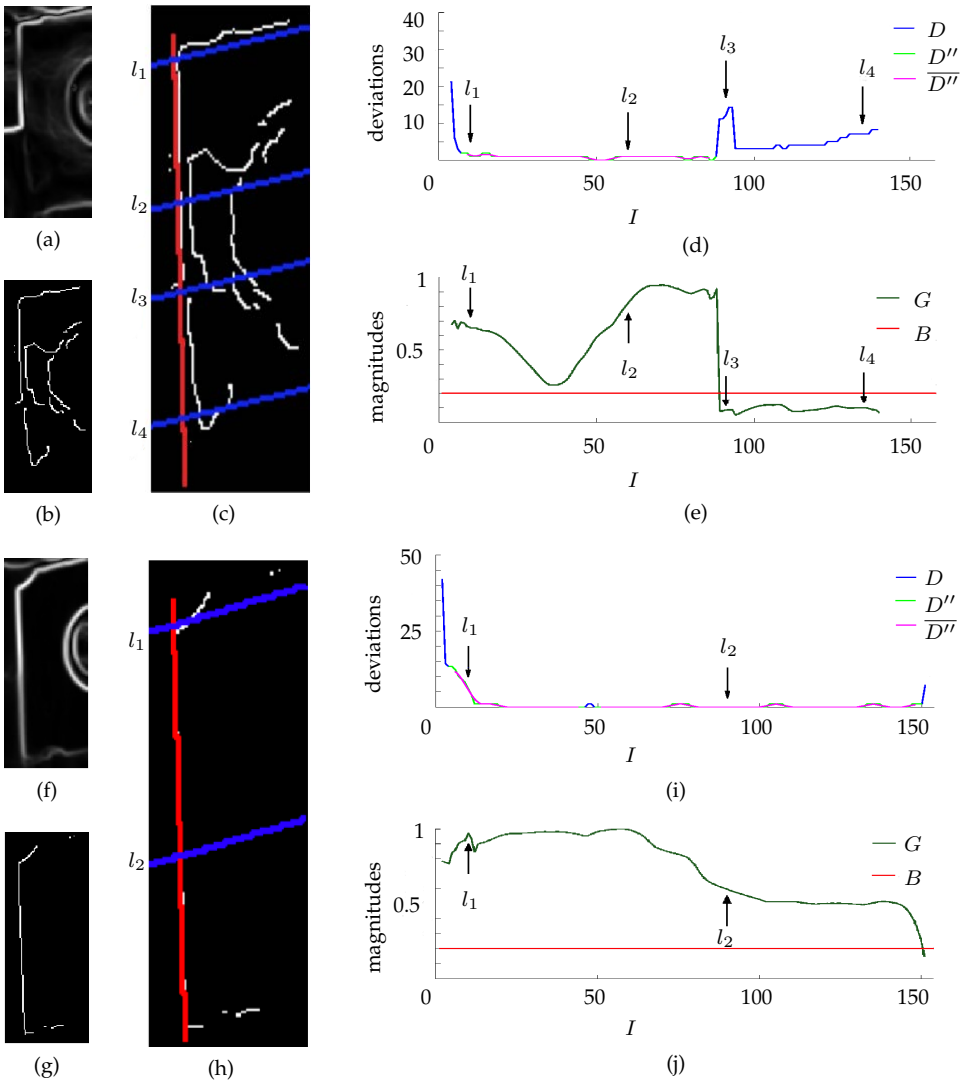


Figure 4.14: Two examples of deviation and gradient magnitude computation. (a-e) and (f-j) correspond to the second and third rows in Fig. 4.13, respectively. (a and f) Edge maps. (b and g) Gradient magnitude maps. (c and h) In white the edge maps, overlapped in red the detected ideal cutting edge and in blue examples of analysed lines $\hat{y}$ as in Eq. 4.3. (d and i) In blue deviations along the detected ideal cutting edge D, in green deviations after spike and low contrast elimination D'' and in magenta deviations after mean filtering $\overline{D''}$. (e-h) In dark green gradient magnitudes along the detected ideal cutting edge and in red an example of threshold $B = 0.2$.

by $f_{A,N}(D,t)$ a function that evaluates a neighbourhood of width A within the given set D centred at point t:

$$f_{A,N}(D,t) = \begin{cases} D_t & \text{if } D_t \leq N \times \text{median}_{j=-A}^{A}(D_{t+j}) \\ \emptyset & \text{otherwise} \end{cases} \tag{4.7}$$

This function returns the element D_t if D_t is higher than a fraction N of the median within a local window of width A, otherwise it returns $\emptyset$. We define a set D' which is formed by applying the function f two consecutive times in order to remove spikes with a length of at most 3 points.

$$D' = \{f_{A_2,N_2}(f_{A_1,N_1}(D,t),t) \mid \forall\, t \in D\} \tag{4.8}$$

The first insert in Fig. 4.14 shows a typical problem in this application. The lower part of its cutting edge has low contrast and as a result the corresponding edge points have very low gradient magnitudes. We are only interested in evaluating the parts along the cutting edge that have high contrast because they are more reliable. Formally, we define a new set D'' whose elements are copied from the set D' when the corresponding edge points have gradient magnitudes higher than a threshold B, otherwise they are set to $\emptyset$.

$$D'' = \{g_t \geq B \rightarrow d_t \wedge g_t < B \rightarrow \varnothing \,|\, d_t \in D',\ \forall\, g_t \in G\} \tag{4.9}$$

In order to ensure that an insert is broken, the deviation should be sufficiently high along a region of the cutting edge and not just in one isolated pixel. We apply a mean filter of window with a width C and subsequently take the maximum deviation $\overline{d}$ of the cutting edge.

$$\overline{d} = \max \left\{ \frac{1}{2C+1} \sum_{j=-C}^{C} (d_{t+j}) \,\middle|\, \forall\, d_t \in D'' \right\} \tag{4.10}$$

Moreover, we also compute the mean gradient magnitude $\overline{g}$ along the cutting edge.

$$\overline{g} = \frac{1}{|G|} \sum_{t=1}^{|G|} g_t,\ \forall\, g_t \in G \tag{4.11}$$

As a result every localised insert is represent by the two parameter values $\bar{d}$ and $\bar{g}$.

Classification of inserts

We remind the reader that the same insert is detected in several images under different poses. In this work, the correspondences of the same insert in multiple images is manually labelled. In the Section 4.3.3 we provide a suggestion how the correspondence issue can be implemented automatically. For each insert we compute the maximum deviation $\bar{d}$ and the mean gradient $\bar{g}$ for every image where it is detected.

We classify an insert as broken if the image with the highest mean gradient magnitude $\bar{g}$ along the cutting edge has a maximum deviation $\bar{d}$ higher than a threshold T, or if the maximum deviations of at least two images (irrespective of the mean gradient magnitude) are greater than T. Otherwise we classify the insert as unbroken. Formally, we define the classification function z as:

$$z(e) = \begin{cases} 1 & \text{if } (\bar{d}_{\text{argmax}_{h=1\dots r}\{\bar{g}_h\}} > T) \vee \left(\sum_{h=1}^{r} \left(\bar{d}_h > T\right)\right) \geq 2 \\ 0 & \text{otherwise} \end{cases} \tag{4.12}$$

where r is the number of images where the same insert e is detected.

4.3.2. Experiments and results

We used Matlab in a personal computer with a 2GHz processor and 8GB RAM. The complete process to identify broken inserts in a head tool with 30 inserts takes less than 3 minutes. This is sufficient for the application at hand because according to the consulted experts the milling tool head stays in a resting position between 5 and 30 minutes, during which the milled plate is replaced by a new one.

Our dataset is skewed with 19 broken inserts and 161 unbroken ones. We refer to the broken inserts as the positive class and the unbroken as the negative class. Therefore, a true positive (TP) is a broken insert classified as broken; a false positive (FP) is an unbroken insert classified as broken and a false negative (FN) is a broken insert classified as unbroken. We compute the precision $P = TP/(TP + FP)$, recall $R = TP/(TP + FN)$ and their harmonic mean $F = 2PR/(P + R)$ for a set of thresholds $T \in \{5, 5.01, \dots, 8\}$ used in the classification function, and obtain a $P-R$ curve. We consider the best pair (P, R), the one that contributes to the maximum harmonic mean.

We apply a repeated random sub-sampling validation where in each run we randomly (stratified sampling) split the dataset into training (70%) and validation (30%) sub sets. For each such split, we use the training data to determine the set of parameters $(A_1, N_1, A_2, N_2, B, C)$ that achieves the global maximum harmonic mean F. This is obtained by applying a grid search on $A_1 \in \{3, 5, 7\}$, $N_1 \in \{1, 1.5, 2\}$,

$A_2 \in \{3, 5, 7\}$, $N_2 \in \{1, 1.25, 1.5\}$, $B \in \{0.18, 0.2, 0.22\}$ and $C \in \{3, 5, 7\}$ and computing the maximum harmonic mean for each combination. If several combinations of parameters yield the same harmonic mean, we take a random one. The determined set of parameters is then used to evaluate the validation set. We repeat this process 20 times and finally we average the results obtained from the validation sets. We obtain an average harmonic mean $F = 0.9143(\pm 0.079)$ with a precision $P = 0.9661(\pm 0.073)$ and a recall $R = 0.8821(\pm 0.134)$. The most repeated (6 out of 20 runs) set of parameters in the training is ($A_1 = 5, N_1 = 1.5, A_2 = 3, N_2 = 1, B = 5, C = 0.2$). When we evaluate the entire dataset with these parameter values we achieve precision $P = 1$ and recall $R = 0.95$ for the maximum harmonic mean $F = 0.9744$.

4.3.3. Discussion

We performed an effective classification of the inserts according to the state of their cutting edges as broken and unbroken. The high performance results that we achieved demonstrate the effectiveness of the proposed approach and suggest that this system can be applied in production. The performance can be further improved by using more appropriate illumination conditions and better quality of the lenses in order to obtain higher contrast between inserts and background.

Typically, an insert appears in 7 to 10 images in different positions and poses. In this work, the ground truth contains the identification numbers of the inserts in all images. This means, that an insert that appears multiple times is manually given the same identification number. Alternatively, the approximate position of inserts in the consecutive images can be inferred from the radius of the head tool cylinder, the distance of the fixed camera from head tool and the degrees of rotation. In this way, after automatically detecting the positions of inserts we can automatically determine the correspondences (labelling) according to the expected positions.

In this work we are concerned with detecting broken inserts as it is the most critical evaluation for the stability of the milling head tool. In future, we will also evaluate the wear of inserts in order to detect the weak ones as early as possible. Moreover, we would also like to compare the performance of different image acquisition methods.

In addition, the proposed methodology can be set up for different machining heads that contain polygonal inserts fastened by screws, a typical design in milling machines.

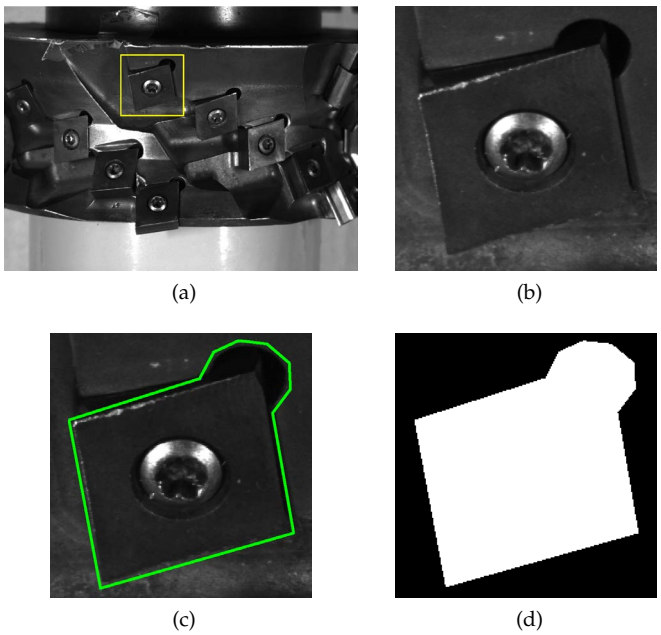
4.4. Automatic localisation of inserts using COSFIRE

In this section we propose a method for the localisation of inserts based on COSFIRE filters. This approach considers independently each image of the dataset and can be automatically configured regardless of the appearance of the inserts. This trainable approach is more versatile and generic than previous works on the topic, as it is not based on, and for that reason does not require, any a priori domain knowledge.

4.4.1. Method

Overview

In order to detect a particular object in an image, COSFIRE filters are first configured by using some training patterns, also referred to as prototypes. We obtain a prototype pattern by extracting a delimited area —region of interest (ROI)— containing one of the inserts in a representative image, Fig. 4.15.



(a) (b) (c) (d)

Figure 4.15: Selection of the region of interest: (a) Input image. (b) Prototypical insert. (c) Selection of the ROI. (d) Mask that mark out the area of the ROI.

COSFIRE filters, Azzopardi and Petkov (2013c), combine the responses of 2D Gabor filters at specific locations around a given point. Gabor filters, Petkov and Wieling (2008), are configured by establishing their characteristic directions and the locations at which their responses are taken. Consequently, the resulting COSFIRE filter only responds to inserts similar in local spatial arrangement to that in the ROI. In this case, the most characteristic edges are found on the sides of the insert, around the screw and on the top right crack, Fig. 4.15.

Gabor filters

The real Gabor function $h_{\lambda,\theta}(x,y)$ for a given wavelength λ and orientation θ is defined as:

$$h_{\lambda,\theta}(x,y) = e^{\left(-\frac{u^2+\gamma^2 v^2}{2\sigma^2}\right)} \cos\left(2\pi\frac{u}{\lambda}+\zeta\right) \quad (4.13)$$

$$u = x\cos(\theta) + y\sin(\theta) \quad (4.14)$$

$$v = -x\sin(\theta) + y\cos(\theta) \quad (4.15)$$

where $\gamma = 0.3$ is the aspect ratio that specifies the ellipticity of the support of the Gabor function; σ determines the size of the support; and $\zeta = \pi/2$ is the phase offset that determines the symmetric or antisymmetric shape of the Gabor function[3].

We denote by $g_{\lambda,\theta}(x,y)$ the response of a Gabor filter to a grayscale input image I:

$$g_{\lambda,\theta}(x,y) = I * h_{\lambda,\theta}(x,y) \quad (4.16)$$

Gabor functions are normalized in such a way so all positive values sum up to 1 whereas all negative values sum up to -1. In this way, the response to an image of constant intensity is 0 even for symmetrical filters ($\zeta = \{0,\pi\}$) and the largest response to a line of width w is achieved using a symmetrical filter ($\zeta = \{0,\pi\}$) with $\lambda = 2w$.

Configuration of COSFIRE filters

A COSFIRE filter is configured by determining the geometrical properties of the lines and edges in the neighbourhood of a specified point of interest, which in this case is the centre of a screw. The neighbourhood is defined by a set of circles of given

[3]For more details about Gabor filters and the use of their parameters such as the aspect ratio or the standard deviation of the Gaussian envelope, we refer the reader to Grigorescu et al. (2003b, 2002); Kruizinga and Petkov (1999); Petkov (1995); Petkov and Kruizinga (1997); Petkov and Westenberg (2003).

Table 4.1: Set of tuples that describe the contour parts of the prototype shown in Fig. 4.15 for a circle with radius $\rho = 107$.

i	λ_i	θ_i	ρ_i	ϕ_i
1	6	$15\pi/8$	107	0.7
2	6	$5\pi/8$	107	1.22
3	6	π	107	4.08

radii. We first superimpose the responses of a bank of Gabor filters with one scale ($\lambda = 6$) and 16 orientations ($\theta = \{0, \pi/8, \dots\}$). For each local maximum Gabor response along these circles we consider the Gabor filters that give a response greater than a fraction t_2 of the maximum Gabor response at that position. Then, we create a 4-tuple $(\lambda, \theta, \rho, \phi)$ for every Gabor filter that satisfies the mentioned criteria: the wavelength λ and orientation θ of the Gabor filter define the characteristics of the concerned Gabor filter, while the distance ρ and polar angle ϕ define the position with respect to the center.

We denote by S_f a COSFIRE filter with a set of 4-tuples $(\lambda_i, \theta_i, \rho_i, \phi_i)$ that characterize the properties of contour parts:

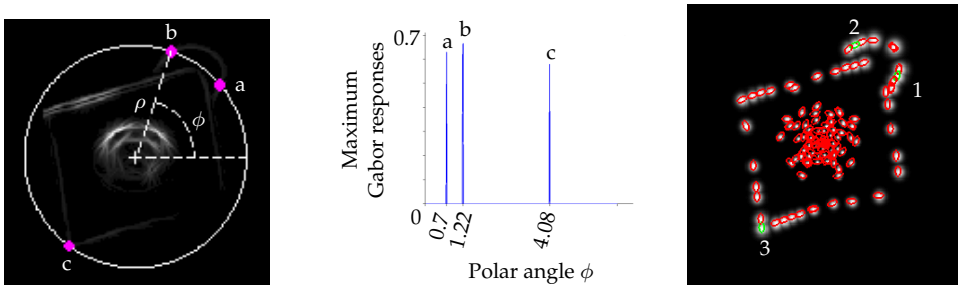
$$S_f = \{(\lambda_i, \theta_i, \rho_i, \phi_i) \mid i = 1, \cdots, n_f\} \tag{4.17}$$

The subscript f stands for the feature (in this case an insert) around the point of interest (ROI) and n_f stands for the number of involved contour parts.

For the ROI shown in Fig. 4.15, taking 25 equally spaced radii from 0 to 150 (the half diagonal of the ROI), this method results in a COSFIRE filter with 127 tuples. Fig. 4.16 illustrates the consideration of Gabor responses for the circle with radius ρ = 107. Along this circle the automatic configuration determines 3 tuples; one for each point a, b and c with parameter values specified in the set shown in Table 4.1. The third tuple $(\lambda_3 = 6, \theta_3 = \pi, \rho_3 = 107, \phi_3 = 4.08)$ describes a contour part with a wavelength of $(\lambda_3 = 6)$ and an orientation of $\theta_3 = \pi$, therefore it is a vertical contour part, that can be detected by a Gabor filter with preferred wavelength $\lambda_3 = 6$ and orientation $\theta_3 = \pi$, at a position of $\rho_3 = 107$ pixels to the bottom-left ($\phi_3 = 1.29\pi$) from the support center of the filter. This location is marked by the label 'c' in Fig. 4.16a. In Fig. 4.16c we illustrate the structure of the resulting COSFIRE filter with 127 tuples.

Application of COSFIRE filters to milling head images

A COSFIRE filter is applied by computing the Gabor filters defined in the set of tuples. Then, for each position in an image, we combine the Gabor responses whose



(a) (b) (c)

Figure 4.16: Configuration of a COSFIRE filter: (a) Superposition of the response maps of a bank of Gabor filters. The white cross indicates the point of interest and the white circle represents the locations of the Gabor responses considered around the point of interest for a given radius ρ, here $\rho = 107$. The gray-level of a pixel represents the maximum value superposition of the responses of a bank of symmetric Gabor filters ($\lambda = 6, \theta = \left\{\frac{\pi i}{8}, i = 0...7\right\}$ and $\zeta = \pi/2$) at that position. (b) The maximum Gabor responses along the depicted circle in (a). The three local maxima in the plot are respectively labelled and marked with magenta dots in (a). (c) Structure of the COSFIRE filter. Each of the ellipses represent a tuple of the set of contour parts. Their size and orientation represent the scale λ and orientation θ parameters of the Gabor filters. This filter is configured to detect the spatial local arrangement of 127 contour parts. The green enumerated ellipses represent the three contour parts found for $\rho = 107$ described in Table 4.1: ellipse 1 corresponds to the local maximum a, ellipses 2 to b and ellipse 3 to c. The bright blobs are intensity maps of the Gaussian functions that are used in the application step for blurring the responses of the Gabor filters. The blurring step is explained in more detail in Section 4.4.1.4.

locations are specified by the polar coordinates in the set of tuples, and combine them with a multivariate output function.

Blurring and shifting. Before computing the output function of a COSFIRE filter, we first blur the Gabor responses in order to allow for some spatial tolerance of the involved contour parts. The blurring consists of a convolution of the Gabor responses with a rotationally symmetric Gaussian lowpass filter $G_\sigma(x, y)$ with standard deviation σ. The standard deviation is a linear function of the distance ρ from the centre of the COSFIRE filter:

$$\sigma = \sigma_0 + \alpha\rho \tag{4.18}$$

We use $\sigma_0 = 0.67$ and $\alpha = 0.04$. The visual system of the brain inspired the choice of the linear function in Eq. 4.18 as explained in (Azzopardi and Petkov, 2013c). The blurred response for the tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ is defined as:

$$b_{\lambda_i,\theta_i,\rho_i}(x, y) = g_{\lambda_i,\theta_i}(x, y) * G_{\sigma_i}(x, y) \tag{4.19}$$

Instead of retrieving the Gabor responses using the polar coordinates specified in tuples of the filter with respect to each pixel in the image, we shift the blurred responses of each Gabor filter by a distance of ρ_i in the opposite direction to ϕ_i. In polar coordinates, we can express this as $(\rho_i, \phi_i + \pi)$, whereas in Cartesian coordinates it is described as an increment $(\Delta x_i, \Delta y_i)$ where $\Delta x_i = -\rho_i \cos\phi_i$ and $\Delta y_i = -\rho_i \sin\phi_i$. We denote by $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ the blurred and shifted response of the Gabor filter specified by the tuple $(\lambda_i, \theta_i, \rho_i, \phi_i)$ in the set S_f:

$$s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) = b_{\lambda_i,\theta_i,\rho_i}(x - \Delta x_i, y - \Delta y_i) \tag{4.20}$$

where $-3\sigma \leq x, y \leq 3\sigma$.

Response of a COSFIRE filter. In the work published in Azzopardi and Petkov (2013c) the response of a COSFIRE filter is defined as the geometric mean of all blurred and shifted responses of the involved Gabor filters as defined in Eq. 4.21. This is a hard AND-type function as the absence of only one of the preferred contour parts suppresses completely the response of the COSFIRE filter, onwards named Hard Geometric Mean (HGM). Here, we experiment with two other softer output functions, namely Arithmetic Mean (AM) and Soft Geometric Mean (SGM), defined in Eq. 4.22 and Eq. 4.23, respectively.

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \left(\prod_{i=1}^{|S_f|} (s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)) \right)^{1/|S_f|} \right|_{t_3} \tag{4.21}$$

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \frac{1}{|S_f|} \left(\sum_{i=1}^{|S_f|} (s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)) \right) \right|_{t_3} \tag{4.22}$$

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \left(\prod_{i=1}^{|S_f|} (s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) + \epsilon) \right)^{1/|S_f|} \right|_{t_3} \tag{4.23}$$

where $|.|_{t_3}$ means that the response is thresholded at a fraction t_3 of the maximum across all coordinates (x,y). The parameter ϵ in Eq. 4.23 is a very small value in order to avoid complete suppression by non-present contour parts. In this work, we set $\epsilon = 10^{-6}$. In this way a COSFIRE filter that uses an SGM output function always gives a response greater than zero. As for the AM metric, the lack of presence of a contour part has a lower effect in the response of the COSFIRE filter than SGM or HGM.

From the COSFIRE response map $r_{S_f}(x,y)$, we first choose the local maxima points by considering neighbourhoods of 8 pixels. Then, if two local maxima points

are within a Euclidean distance of 200 pixels, we only keep the point with the strongest response. Due to the shape of the milling cutting head and the conditions of the image capture, inserts are always separated by at least 200 pixels. We call these points, positive response points.

Figure 4.17 shows the whole process of edges detection. In this example, the COSFIRE filter is applied with the SGM function and has 127 tuples. Each blurred and shifted response corresponds to each of the 127 contour parts found in the configuration. The filter responds in locations where there is an identical or similar pattern to the prototypical insert. In this example, the maximum response is reached in the center of the prototype insert that was used to configure this COSFIRE filter and the other four local maxima points correspond to inserts that are similar to the prototypical insert.

4.4.2. Experiments

The dataset is split in two subsets, training and test. The training set is formed by the images of the dataset separated by 13 snapshots with numbers 0001, 0014, 0028, 0042, 0056, 0070, 0084, 0098, 0112 and 0126. The other 134 images form the test set.

We configure filters in an iterative process by using inserts from the training images. We configure a filter S_{f_1} for prototype f_1, shown in Fig. 4.18a. Then, we apply this filter to all the images in the training set. We set the value of t_3 to produce the highest number of correctly detected inserts and no false positives, therefore achieving 100% precision. Figure 4.19 shows the inserts found with functions AM, HGM and SGM using the filter for prototype f_1. Threshold t_3 is set to 0.283, 0.044 and 0.119 for AM, HGM and SGM detecting 9, 35 and 37 correct inserts respectively. In total, there are 86 inserts in the 10 training images. Thus, this single COSFIRE filter detects 43.02% of the inserts using SGM, and no false positives.

In the second iteration, we randomly choose one of the inserts that was not detected by the first filter S_{f_1} and we call it prototype f_2. We use this prototype to configure a second COSFIRE filter S_{f_2}. Then, we apply this filter to the 10 images of the training set and determine the t_3 parameter values that achieve 100% precision. Filter S_{f_2} detects an amount of inserts, some already detected by filter S_{f_1} and some new detections. For example, S_{f_2} with SGM correctly detects 14 inserts, of which 4 coincide with the inserts detected by S_{f_1} and 10 are newly detected ones. At this point, we have detected a total of 47 inserts out of 86.

The process successively continues until all the 86 inserts in the training set are detected. We configure a total of 19 filters for HGM, from the prototypes shown in Fig. 4.18, for yielding 100% precision at 100% recall, only the first 17 filters are

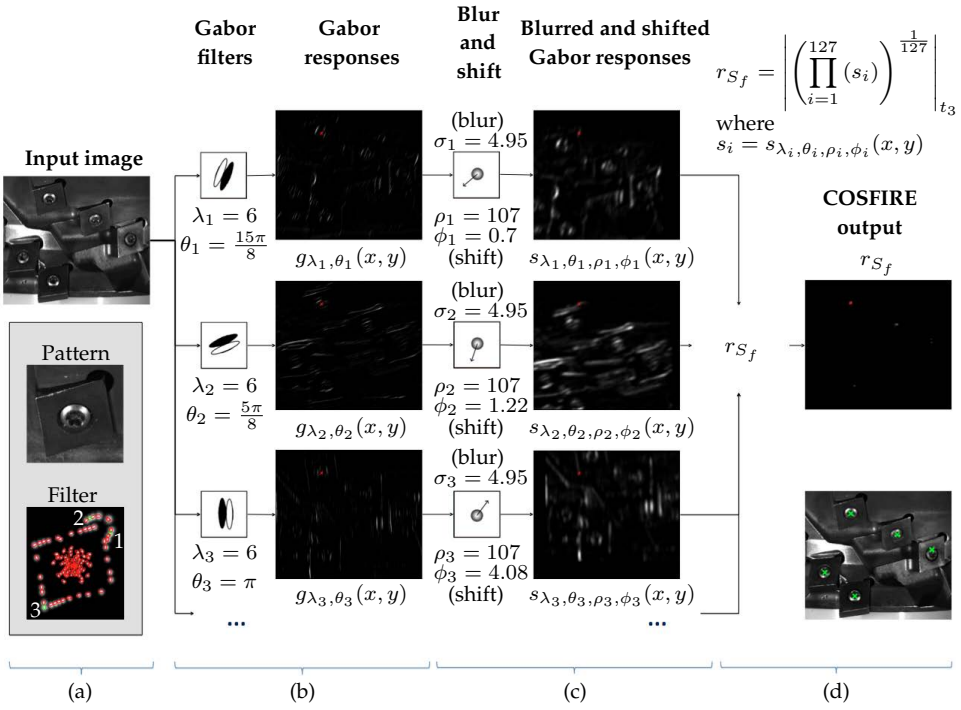
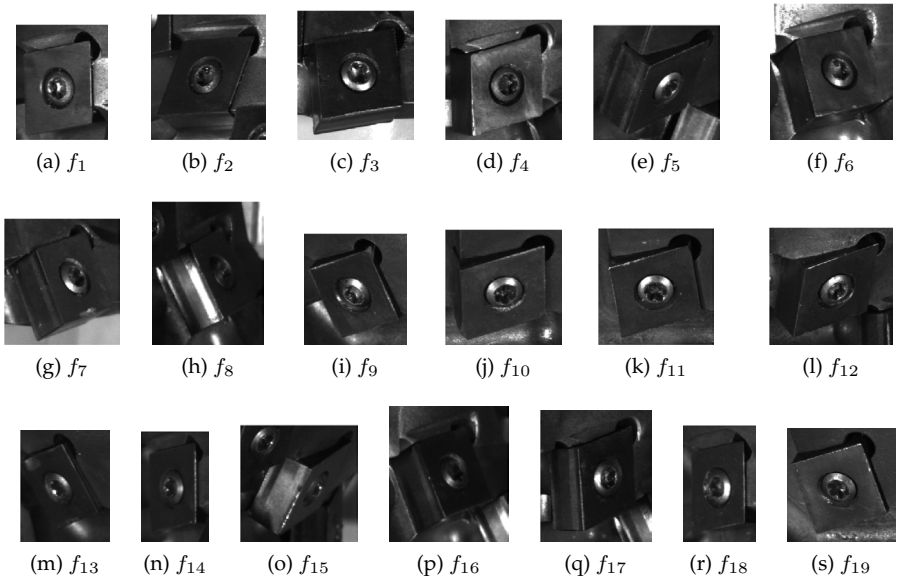


Figure 4.17: (a) Input image. We show just part of the input image for better visualization. The framed area shows (top) the enlarged pattern of interest selected for the configuration and (bottom) the structure of the COSFIRE filter that was configured for this pattern. The contour parts found at $\rho = 107$ whose application is shown in this figure are numbered and marked in green color. (b) Each contour part of the prototype pattern is detected by the response of an antisymmetric Gabor filter with preferred values of wavelength λ_i and orientation θ_i. In this case, we need a Gabor filter to detect each of the contour parts. In general, contour parts with the same pair of values (λ_i, θ_i) are detected by the same Gabor filter. (c) The response $g_{\lambda_i,\theta_i}(x,y)$ is then blurred and later shifted by $(\rho_i \phi_i + \pi)$ in polar coordinates. (d) Finally, the output of the COSFIRE filter is computed by the thresholded soft geometric mean of all the contour part responses, for this example $t_3 = 0.15$. The five local maxima in the output of the COSFIRE filter correspond to the configured insert and four other similar inserts in the input image. The red '$\times$' marker indicates the location of the specified point of interest.

necessary when using SGM. The number of filters needed for each output function are reported in Table 4.2.

The set of configured COSFIRE filters is applied to the test set were results are computed in terms of precision, recall and and their harmonic mean, also known as F-Score:



(a) f_1 (b) f_2 (c) f_3 (d) f_4 (e) f_5 (f) f_6

(g) f_7 (h) f_8 (i) f_9 (j) f_{10} (k) f_{11} (l) f_{12}

(m) f_{13} (n) f_{14} (o) f_{15} (p) f_{16} (q) f_{17} (r) f_{18} (s) f_{19}

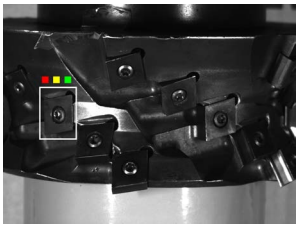
Figure 4.18: A set of 19 prototypical inserts. The whole set was needed to detect all inserts of the training set with 100% precision and 100% recall with HGM function. The first 17 filters were needed when using SGM.

$$\mathrm{F}_{\mathrm{Score}} = 2\frac{Precision \cdot Recall}{Precision + Recall} \tag{4.24}$$

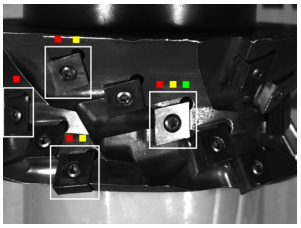
Recall is the percentage of true inserts that are successfully detected, $Recall = TP/(TP + FN)$. Precision is the percentage of correctly detected inserts from all positive response points, $Precision = TP/(TP + FP)$. TP, FP and FN stand for true positives, false positives and false negatives, respectively.

4.4.3. Results

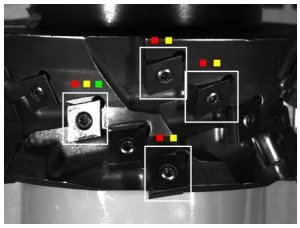
We evaluated the performance of the detection of inserts by a set of COSFIRE filters and we compared results using different output functions. Results are shown in Table 4.2. With AM, 24 COSFIRE filters were configured and applied to the test set yielding an F-Score of 79.83%. A set of 19 filters was configured for HGM reaching a F-Score of 89.76%. SGM required only 17 filters and it achieved 88.89% F-Score. We can conclude that the output functions based on geometric mean are more appropriate than arithmetic mean for detecting inserts.



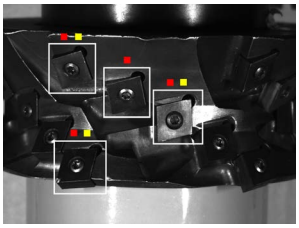
(a) HeadTool0001.bmp



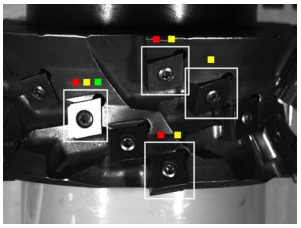
(b) HeadTool0014.bmp



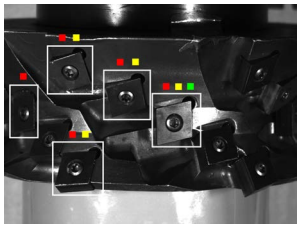
(c) HeadTool0028.bmp



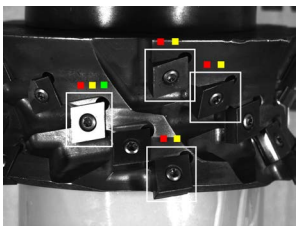
(d) HeadTool0042.bmp



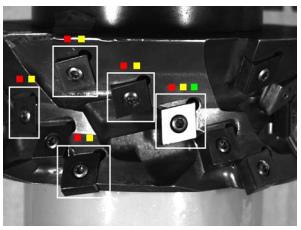
(e) HeadTool0056.bmp



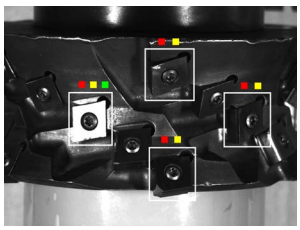
(f) HeadTool0070.bmp



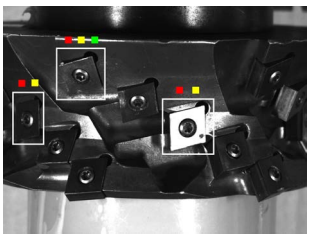
(g) HeadTool0084.bmp



(h) HeadTool0098.bmp



(i) HeadTool0112.bmp



(j) HeadTool0126.bmp

Figure 4.19: Results of applying the filter configured for prototype f_1 to the training set. Detected inserts are marked with a white rectangle. Above the rectangle, a colored square indicates by which output function the insert was found. Red denotes SGM, yellow HGM and green AM.

Table 4.2: Results in terms of number of configured COSFIRE filters, precision, recall and F-Score for the different output functions evaluated: Aritmetic Mean (AM), Hard Geometric Mean (HGM), Soft Geometric Mean (SGM), SGM when configuring the same 19 COSFIRE filters than for HGM (SGM_{19}).

	AM	HGM	SGM	SGM_{19}
number of filters	24	19	17	19
Precision (%)	81.77	**92.62**	92.25	92.39
Recall (%)	78.03	87.08	85.76	**87.52**
F-Score (%)	79.83	89.76	88.89	**89.89**

Besides, the number of configured filters affects detection rates as shown in Azzopardi and Petkov (2013a). They proved that the performance results change with a different number of such filters, for their application, harmonic mean increased when increasing the number of configured filters up until 6 and then it progressively decreased. In order to compare the output functions SGM and HGM, we used the 19 COSFIRE filters that were configured with the HGM method and applied them with the SGM output function. In this experiment we obtained an F-Score of 89.89%, which is better than the F-Score of 89.76% (improvement of 1.12%) that we achieved with the same 19 filters but using the HGM function.

Although COSFIRE filters can achieve tolerance to rotation, scale and reflection Azzopardi and Petkov (2013c), in this application we did not apply any invariances to such geometrical transformations.

Changing the values of the parameter t_3 reaches different performance results. Increasing the value of t_3 causes an increase of precision and a decrease of recall. For each COSFIRE filter, we added to (or subtracted from) the corresponding learned threshold value t_3 an offset value in steps of $0.01t_1$. For all the studied function outputs, the maximum F-Score was reached at values of the threshold parameter t_3 with 0 offset (Fig. 4.20). Thus, the configured values of threshold t_3 at the training set are proven to be the best threshold values also for the test set.

4.4.4. Discussion

In the literature of machine vision, there are three families of approaches that are typically used for the detection of patterns of interest in images.

The first family of solutions are methods based on keypoint descriptors, such as SIFT Lowe (2004), SURF Bay et al. (2008), HOG Dalal and Triggs (2005b), CCS Jacobson et al. (2007). We attempted to use that approach for our application (data not shown), but resulted in lower performance. It is our belief that the reason for

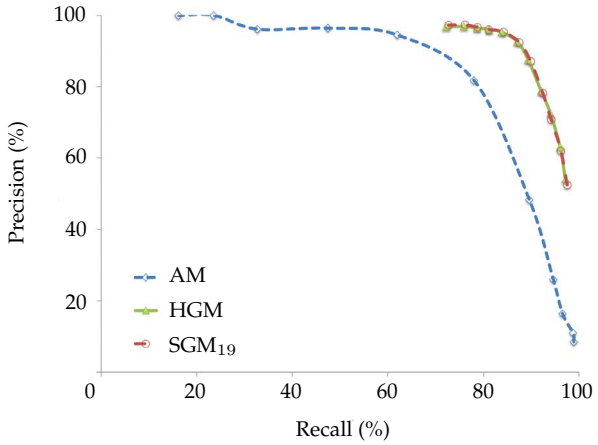


Figure 4.20: Precision-recall curves obtained for each of the studied metric functions: Aritmetic Mean (AM), Hard Geometric Mean (HGM), SGM when configuring the same 19 COSFIRE filters than for HGM (SGM_{19}) and HGM with rotation invariance (HGMr). For each plot, the threshold values of parameters t_3 are varied by adding the same proportional offset value, ranging between $-0.05t_3$ to $0.05t_3$ in intervals of $0.01t_3$, to the corresponding learned threshold values. Precision increases and recall decreases with an increasing offset value. The F-Score reaches the maximum values for each plot at the original offset value t_3 with 0 offset.

this is that, in our case of study, the information lies within the shape and contour of the object, rather than in its texture. Methods based on keypoint descriptors are more suitable for textured surfaces.

The second family consists of those methods based on template matching. Template-matching methods use a set of typical image patterns or templates to determine similarities of an inspection image to a particular pattern in order to make classification decisions in automated visual inspection Sun et al. (2012). A previous work of the authors Aller-Álvarez et al. (2015) applied template matching to this problem and obtained lower performance (F-Score=86%, precision 82% and recall 89% on the same dataset) than those obtained with the approach reported in this paper. In that work, first the authors preprocessed the images by applying Canny's algorithm to the input image followed by a dilation of the edge map with a flat diamond-shaped structuring element of size 1 pixel from the centre of the structuring element to the points of the diamond. Then, they performed a normalized cross-correlation to measure the correspondence between each template, manually selected by the user and the considered window in the input image. The response of the template matching was considered as the two best correspondences per input image and template. The same test and training sets as in this work were used for

obtaining the experimental results.

The third family of solutions are those that use domain knowledge. For instance, in this particular application we know that an insert is made of a circular screw surrounded by a rhomboid shape. We attempted this approach and obtained good results Fernández-Robles et al. (2015). In that case, the presence of a screw allowed the identification of the insert by means of detecting its circular contour.

The approach reported in the present section is far more versatile as it can also be applied to identify any tool or part without using domain knowledge. This is particularly important in other machine vision applications with objects of interests that might be very different than the inserts in the concerned application.

4.5. Conclusions

The contributions of the work presented in this chapter are four-fold. First, we described a method for the localisation of inserts with independence among images. Second, the approach that we proposed for the localisation of cutting edges in milling machines is highly effective and efficient. Its output is a set of regions surrounding cutting edges, which can be used as input to other methods that perform quality assessment of the edges. Third, we achieved an effective classification of the inserts with respect to the state of their cutting edges as broken and unbroken. Fourth, we presented a dataset of 144 images of a rotating edge milling cutting head that contains 30 inserts, analysing 180 inserts in total. It contains the ground truth information about the locations of the cutting edges, the locations of the centres of the inserts and broken inserts are labelled by experts. We made our dataset publicly available[2].

To our knowledge, this is the first automatic solution for the identification of broken inserts in edge profile milling heads. The presented system can be set up on-line and it can be applied while the milling head is in a resting position without delaying any machining operations. This system highly reduces the risk of head tool collapse, which is very expensive and time consuming to replace.

Chapter 5

Object recognition for content-based image retrieval

5.1. Evaluation of clustering configurations of SIFT features for object recognition applied to CBIR

In this section we evaluate different techniques to determine when there is a correspondence between images and to compute the strength of the correspondence. On the one hand we use the similarity of the closest pair of keypoint descriptors. On the other hand we use a Hough transform to identify clusters of at least three points voting for the same pose of an object and we verify the consistency of the pose parameters with the least squares algorithm. We use different values for the Hough transform parametrization.

5.1.1. Method

We obtain SIFT keypoints and descriptors for the query object images and for all images of the dataset. Then for each query image, we compute the cosine similarity between a descriptor of the ROI with all descriptors of the query image. For this ROI descriptor, we consider the match that obtains the maximum similarity (minimum cosine angle) as long as its cosine angle is less than 2 times the cosine angle of the second nearest neighbour. Otherwise, we discard that match. Repeating this computation for all descriptors of the ROI, we obtain a set of matches between a ROI and a query image. Afterwards, we either use directly this information or we perform a voting and a geometric verification for pose of the object to decide about the correspondence between images.

On the one hand, we consider the correspondence of the match that achieves the minimum cosine angle among all matches between ROI and query image after the second nearest neighbour test. The pair of keypoints with the smallest angle is the most similar one among all pairs of matched keypoints and therefore this match has the highest probability of being correct. We use the value of such cosine angle of the

most similar pair of keypoints as a measure of the similarity between the ROI and the query image. The hit list is ranked by sorting the retrieved images in ascending order in relation to this metric. We refer to this case as *without clustering*.

On the other hand, from the set of matches between the ROI and the query image we identify clusters of keypoints that vote for the same pose of an object using Hough transform and we perform a geometric verification using least squares algorithm as suggested by Lowe (2004).

Each SIFT keypoint specifies 4 parameters: 2D location, scale and orientation. We keep track of these parameters for the match keypoints. Therefore, we can create a Hough transform entry predicting the model location, orientation, and scale from the match keypoints. The Hough transform creates a four dimension accumulator and uses each keypoint set of parameters to vote for all object poses that are consistent with it. When clusters of keypoints vote for the same pose of an object, it is more probable that they belong to the same object than just relying on a single keypoint (Lowe, 2004). Each keypoint match votes for the 2 closest bins in each dimension to solve the problem of boundary effects in bin assignment. Lowe's clustering uses broad bin sizes of 30 degrees for orientation, a factor of 2 for scale, and 0.25 times the maximum projected training image dimension (using the predicted scale) for location. We refer to this case as *Lowe's clustering*.

Afterwards, we use least squares algorithm to seek for geometric verification. We require each match in a cluster to agree within the Hough model, otherwise we consider that match as an outlier and it is removed. If less than three keypoints remain after discarding outliers, we reject the whole cluster of matches.

Finally for each remaining cluster, we compute the average of the cosine angles of the matches within the cluster. We take the minimum average for all clusters as a measure of the similarity between the ROI and the query image. Again, the hit list of retrieved images is sorted in ascending order according to this metric.

We evaluate other choices of the parameters used in the Hough transform model. We aim at obtaining a less restrictive clustering of matches by broadening their size (so lowering the number of bins). By considering broader bins, more keypoints agree for the same object pose. At the same time, less false correspondences are rejected. *Half* and *quarter clustering* settings use 60 and 90 degrees for orientation, factor of 4 and 6 for scale, and 0.5 and 0.75 times the maximum projected training image dimension for location respectively.

Figure 5.1: Examples of images containing the same object, a blue toy car. Changes in pose, scale, orientation, illumination and cluttered background can be noticed making the object retrieval task very challenging.

5.1.2. Evaluation

dataset

For the purpose of ASASEC, retrieving objects from a dataset containing child pornography, we have created and made public our own dataset[1]. It is composed of 614 frames of 640×480 pixels that come from 3 videos. All videos were recorded in different bedrooms with different distributions, illumination, textures, etc., making the object retrieval a challenging task, Fig. 5.1. Nevertheless some objects are present in all videos such as two toy cars, some clothespins, a stuffed bee, some pens, some cups or a child book together with a big doll. The doll is usually the principal actor in the videos and helps us to simulate partial occlusions of the objects and a more realistic scenario. Although these objects are present in every video, they do not appear in every frame. Together with them, other objects are unique in each bedroom. We also provide a ground truth indicating which objects are visible in each frame.

[1] dataset is available at http://pitia.unileon.es/varp/galleries

Figure 5.2: ROIs of the query objects.

Table 5.1: Description of the query objects. Number of images that contain each query object in the dataset of 614 images. Size of each object ROI in pixels.

Object	**Number of query objects**	**Size of the ROI (pixels)**
Book	115	305×334
Blue car	102	285×258
Yellow car	138	208×265
Pink clothespin	125	146×132
Blue clothespin	92	85×145
Green clothespin	42	68×59

5.1.3. Experiments and results

As query objects we have used the book, the blue and yellow car, and the pink, blue and green clothespin shown in Fig. 5.2. The total number of query objects present among the 614 frames of the dataset and the size of the ROIs are specified in Table 5.1.

When dealing with object retrieval, it is important that the retrieved images are ranked according to their relevance to the query object instead of just being returned as a set. The most relevant hits must be in the top few images returned for a query.

Table 5.2: Precision at cuts of the query objects using different clustering parameters. Best results for each precision at n are marked in bold.

	Book					Blue clothespin		
	P@40	P@50	P@60	P@70	P@80	P@5	P@10	P@20
Without	**1**	**1**	**0.9**	0.8	0.75	**1**	**0.7**	**0.35**
Quarter	0.85	0.82	0.77	0.7	0.66	0.4	0.2	0.1
Half	0.93	0.88	0.88	**0.83**	**0.8**	0.4	0.2	0.1
Lowe's	**1**	0.96	0.85	**0.83**	0.79	0.8	0.4	0.2
	Blue car					Pink clothespin		
	P@5	P@10	P@20	P@30	P@40	P@5	P@10	P@20
Without	**1**	**1**	0.75	0.57	0.43	**0.8**	**0.4**	**0.25**
Quarter	1	0.8	0.6	0.43	0.38	0.2	0.1	0.05
Half	0.8	0.9	0.85	0.7	0.55	0.2	0.2	0.1
Lowe's	0.8	0.9	**0.9**	**0.73**	**0.625**	**0.8**	**0.4**	**0.25**
	Yellow car					Green clothespin		
	P@5	P@10	P@20	P@30	P@40	P@5	P@10	P@20
Without	**1**	0.9	0.75	**0.73**	0.63	**1**	0.5	0.3
Quarter	0.8	0.7	0.55	0.47	0.38	0.2	0.1	0.05
Half	**1**	0.8	0.75	0.7	**0.68**	0.8	0.4	0.2
Lowe's	**1**	**1**	**0.85**	0.7	0.65	**1**	**0.7**	**0.35**

Recall and precision are measures for the entire hit list and do not account for the quality of ranking the hits in the hit list. Relevance ranking can be measured by computing precision at different cut-off points, this is technically called precision at n or P@n. Let h $[i]$ be the ith hit in the hit list and let rel $[i]$ be 1 if h $[i]$ is relevant and 0 otherwise. For a hit to be relevant the query object has to be present in the image and correctly localised. Therefore, if the image contains the object but the correspondence is not within that object, rel $[i]$ is 0. Then precision at hit n is:

$$\text{P@}n = \sum_{k=1..n} \text{rel}\,[k]\,/n \tag{5.1}$$

Table 5.2 shows the results for the four clustering types: without clustering, quarter clustering, half clustering and Lowe's clustering showing the precision at different hits.

Examples of the second, fifth and twentieth hit of the hit list for the blue car with the different clustering approaches are shown in Fig. 5.3.

The ROI of the book has well-defined corners that can produce distinctive keypoints easier to detect and match among the images of the dataset. Without clustering approach correctly retrieved the first 51 hits just relying on the strongest match. This

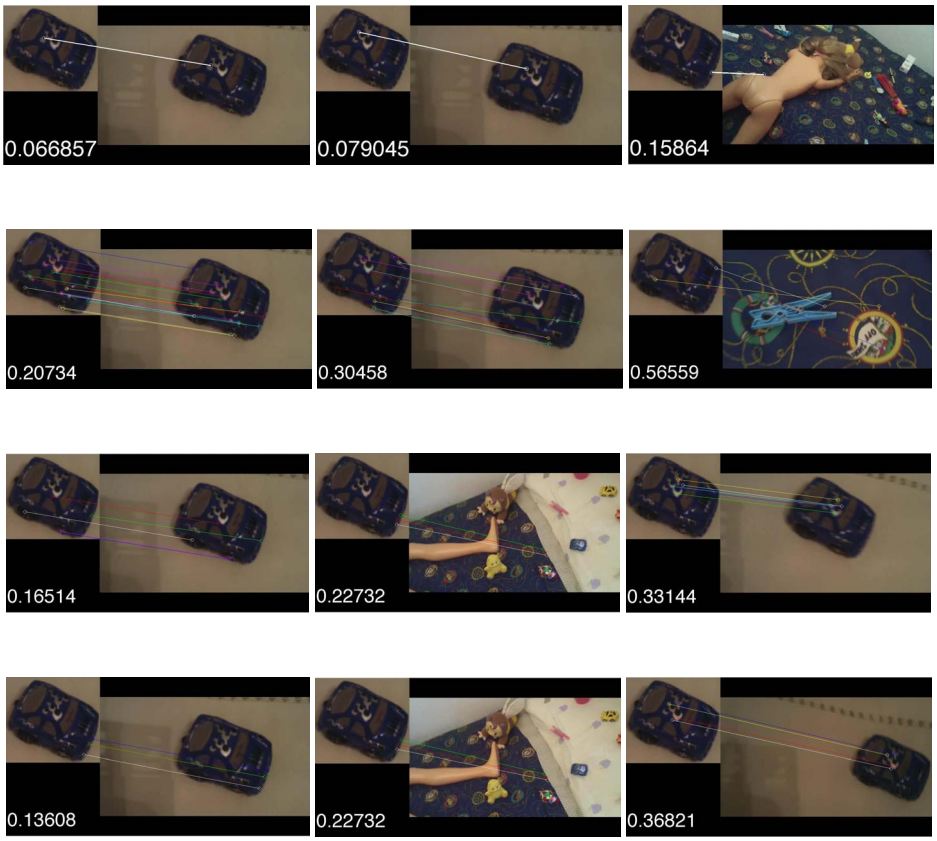


Figure 5.3: Second, fifth and twentieth hits using different clustering parameters for the blue car. In rows: without clustering, quarter clustering, half clustering and Lowe's clustering. In columns: second, fifth and twentieth hit of the hit list. The white value indicates the cosine angle of the match or the average cosine angles of the matches.

is a good result considering that there are 115 images containing the book in all 614 images of the dataset. However for higher cuts in hit list, Lowe's clustering and half clustering approaches obtained higher precision results.

As for the cars, although SIFT method computes the descriptors using gray level images, there are small differences in shape and patterns between the two cars. Regarding the blue car, without clustering yielded the best results for low cuts of the hit list and Lowe's clustering did for high cuts. For the yellow car, without clustering, half clustering a Lowe's clustering obtained similar results.

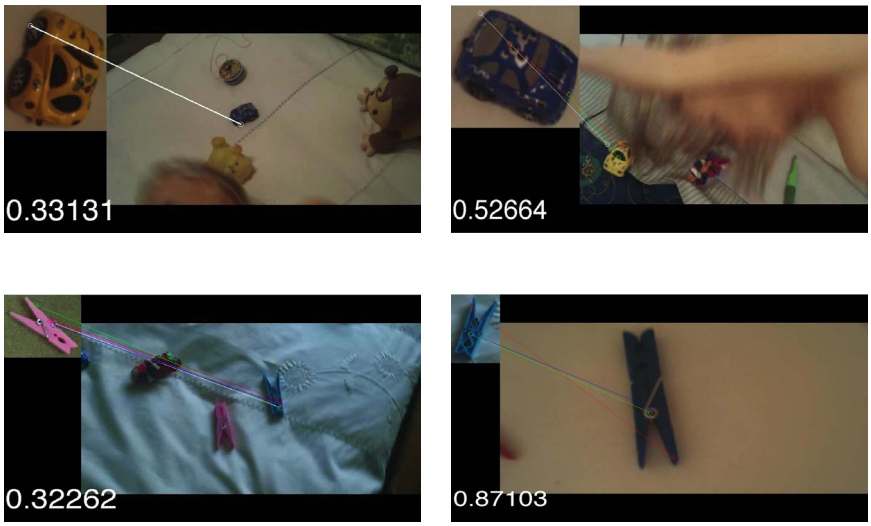


Figure 5.4: Examples of mismatches due to very similar objects that mainly differ in their colours.

Clothespins introduce a more difficult task since less distinctive keypoints are present and their shapes are very alike among them and with some other clothespins of the dataset. Most of the keypoints were found in the metal wire, near the holes or in the outlines. Precision at 20 only reached 0.35, 0.25 and 0.35 for the blue, pink and green clothespins respectively. Without clustering configuration achieved better results for retrieving the blue clothespin, Lowe's clustering performed better for the green one and both approaches obtained the same results for the pink one.

All in all Table 5.2 shows that without clustering approach is more convenient for high precision at small cuts and Lowe's apprach at high cuts of the hit list in this dataset.

In Fig. 5.4 we present examples of misclassified query objects that have a similar shape but different colours leading to mismatches. This is because SIFT is not invariant to colour. Only three mismatches among different cars appeared in all the experiments for the first 20 hits of the hitlist but up to 20 in the case of the clothespins. This could be solved using a colour version of SIFT (Van de Sande et al., 2010).

Background is also another source of mismatches. The pattern duvet of one of the settings leads to many non relevant but distinctive keypoints that locally described can look similar to other objects. For example, some patterns of the duvet are similar to the patterns of the yellow car. Moreover, background of the ROI can

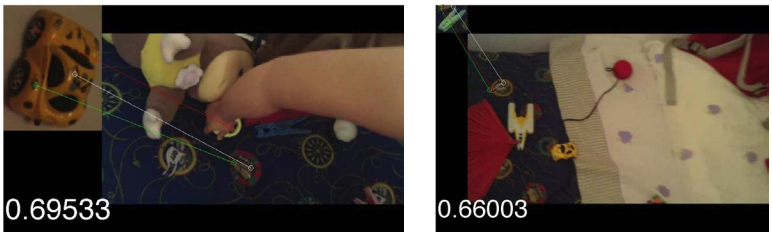


Figure 5.5: Mismatches produced by pattern duvet.

contain distinctive keypoints that produce correspondences with other objects or backgrounds. Figure 5.5 shows examples.

5.2. Adding colour description to COSFIRE filters

In this section we present colour COSFIRE filters. They are trainable keypoint detection operators which are selective for given local colour patterns that consist of combinations of colour contour segments. It is based on COSFIRE filters for gray scale images introduced by Azzopardi and Petkov (2013c). Moreover, colour COSFIRE filters also add invariance to background intensity.

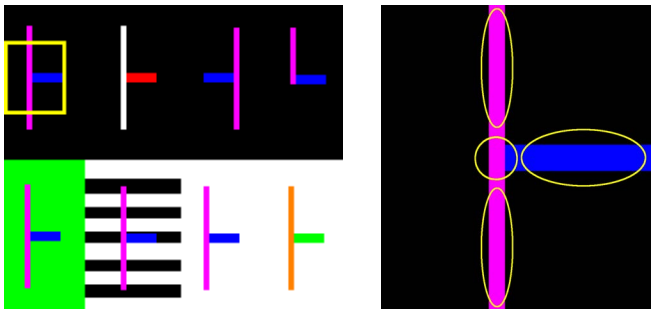
5.2.1. Method with application for colour vertex localisation

Overview

Figure 5.6a shows an input image with 8 vertices. We consider the vertex enclosed in the yellow rectangle as a (prototype) pattern of interest. The rectangular region is known as a region of interest (ROI). This ROI is shown enlarged in Fig. 5.6b. The colour COSFIRE filter configured from this prototype will respond to the same and similar patterns regardless of the background. The prototype has been manually selected by a user.

In Fig. 5.6b, the three ellipses represent the dominant orientations in region of interest. The circle denotes that several dominant orientations are overlapped. We detect the lines by symmetric Gabor filters and describe the colour of these lines by averaging the pixel values of a support region around the centre point of the ellipses for each colour channel.

We compute responses both for line detection and colour description at the centres of the corresponding ellipses in an input image. The response of line de-



(a) (b)

Figure 5.6: (a) Input image of size 180×161 pixels. The yellow square marks the ROI from which the colour COSFIRE filter will be obtained. (b) Enlargement of the ROI. The ellipses represent the support of line detectors that are relevant for the concerned prototype.

tection is computed by applying Gabor filters. The preferred orientations and bandwidths of the Gabor filters and the locations at which we take their responses are automatically determined at the configuration of the colour COSFIRE filter by analysing the prototype pattern. Therefore, the filter only responds to the same (or very similar) local spatial arrangement of lines of specific orientations and widths as in the prototype pattern. We compute the response of colour description of each line segment as the average of the pixel values in a support area around the centres of the corresponding ellipses for each colour channel. We compute the colour response by a Gaussian kernel that measures the similarity between the colour descriptions of the prototype and the input image. Thus, the filter only responds to the same (or very similar) local spatial arrangements of colours as in the prototype pattern.

The response of a colour COSFIRE filter is computed by multiplying the responses of the line detection and the responses of the colour description achieved in the centres of the corresponding ellipses and combining all the multiplications. The response of a colour COSFIRE filter comes from a pixel-wise evaluation of a multivariate function. For that purpose, the responses of Gabor filters and the responses of Gaussian kernels at locations around a pixel are previously shifted to come to that point.

In the following sections we explain the automatic configuration of a colour COSFIRE filter and its application to an input image.

Configuration of a colour COSFIRE filter for vertex localisation

Detection of orientations. We build the colour COSFIRE filter from the responses of 2-Dimensional Gabor filters applied to each colour channel. Gabor filters allow lines or edges detection, depending on their configuration, discriminating frequencies and orientations. Filtering individually the three colour channels and then combining these three responses increases illumination invariance and discriminative power leading to a more accurate detection of the activations in the image than filtering a luminance channel (Van de Sande et al., 2010), as for example the gray level image.

We denote by $g_{\lambda,\theta,\zeta,c}(x,y)$ the response of a Gabor filter of preferred wavelength λ, orientation θ and phase offset ζ to a given colour channel c of the prototype image P. Regarding the considered phase offset of the sinusoidal wave function ζ, the Gabor filter could be symmetric ($\zeta \in \{0,\pi\}$), antisymmetric ($\zeta \in \{\pi/2, 3\pi/2\}$) or an energy filter by taking a quadrature pair of symmetric and antisymmetric phase offsets. For more details about Gabor filters and the use of their parameters (aspect ratio, the standard deviation of the Gaussian envelope, etc.), we refer the reader to (Petkov, 1995; Petkov and Kruizinga, 1997; Kruizinga and Petkov, 1999; Grigorescu et al., 2002; Petkov and Westenberg, 2003; Grigorescu et al., 2003b,a). We normalise the Gabor functions that we use so all positive values sum up to 1 whereas all negative values sum up to -1. In this way, the response to an image of constant intensity is always 0 and the largest response to a line of width w is achieved using a symmetrical filter with $\lambda = 2w$.

The response of a Gabor filter is computed by convolving the input image with a Gabor kernel of preferred parameter values. We obtain a new kernel from each given Gabor kernel that we use. For symmetric filters, the new kernel is made up from the central part of the Gabor kernel whereas for antisymmetric filters it is made up from the largest positive part of the Gabor kernel. We denote by $K_{\lambda,\theta,\zeta}$ such a kernel associated with its corresponding Gabor response $g_{\lambda,\theta,\zeta,c}(x,y)$, note that the same kernel is used for every colour channel.

In order to detect lines or edges, we compute the L-infinity norm of the three Gabor responses obtained for each colour channel.

$$g_{\lambda,\theta,\zeta}(x,y) = \max_{z=1,2,3} \{g_{\lambda,\theta,\zeta,c_z}(x,y)\} \tag{5.2}$$

Then, we compute the L-infinity norm across the two values of ζ used. We use $\zeta = \{0,\pi\}$ for line detection and $\zeta = \{\pi/2, 3\pi/2\}$ for edge detection. We analyse

both values of ζ to achieve independence from the background luminance.

$$g_{\lambda,\theta}(x,y) = \max_{z=1,2} \{g_{\lambda,\theta,\zeta_z}(x,y)\} \tag{5.3}$$

Finally, we threshold the responses of Gabor filters at a fraction t_1 ($0 \leq t_1 \leq 1$) of the maximum response of $g_{\lambda,\theta}(x,y)$ across all combinations of values (λ,θ) used and all positions (x,y) in the image, and denote these thresholded responses as $|g_{\lambda,\theta}(x,y)|_{t_1}$. This operation rejects low responses of Gabor filters that fall under a local threshold.

Contour part and colour description. The colour COSFIRE filter is configured around a selected point of interest, which we consider as the centre of the filter. This point can be either manually selected by an user or automatically set as the central pixel of the ROI. We take the responses of a bank of Gabor filters, characterised by parameter values (λ,θ), along circumferences of given radii ρ around the point of interest, Fig. 5.7. When $\rho = 0$, we only consider the point of interest. The colour COSFIRE filter is defined at certain positions (ρ_i,ϕ_i) with respect to the point of interest in which there are local maxima responses of the bank of Gabor filters. A set of seven parameter values $(\lambda_i,\theta_i,\rho_i,\phi_i,\gamma_{1_i},\gamma_{2_i},\gamma_{3_i})$ characterizes the properties of a contour part that is present in the specified pattern of interest: $\lambda_i/2$ represents the width, θ_i represents the orientation, (ρ_i,ϕ_i) represents the location and $(\gamma_{1_i},\gamma_{2_i},\gamma_{3_i})$ represents the colour description at each colour channel. In the following, we explain how we obtain the parameter values of such contour parts.

First, we consider the responses of a bank of Gabor filters, $|g_{\lambda,\theta}(x,y)|_{t_1}$, along a circumference of radius ρ around the point of interest. In each position along that circumference, we take the maximum of all responses across the possible values of (λ,θ) used in the bank of filters. The locations with the highest local maxima within a neighbourhood along an arc of angle $\pi/8$ define the points that characterise the dominant orientations around the point of interest. We determine the polar coordinates (ρ_i,ϕ_i) of such locations with respect to the point of interest.

For such a location (ρ_i,ϕ_i), we consider all combinations of (λ,θ) for which the corresponding responses $|g_{\lambda,\theta}(x,y)|_{t_1}$ are greater than a fraction $t_2 = 0.75$ of the maximum of $|g_{\lambda,\theta}(x,y)|_{t_1}$ across the different combinations of values (λ,θ) used. For further comment on the choice of the value of t_2, we refer the reader to (Azzopardi and Petkov, 2013c). For each value θ that satisfies the previous condition, we consider a single value of λ, the one for which the corresponding response is the maximum of all responses across all values of λ. Each of the previous pairs (λ,θ) in the location (ρ_i,ϕ_i) describe partly a tuple $(\rho_i,\phi_i,\lambda_i,\theta_i)$.

As for the colour description of the tuples, we compute the average of the pixel values in a region around the location (ρ_i,ϕ_i) for each colour channel. We centre the

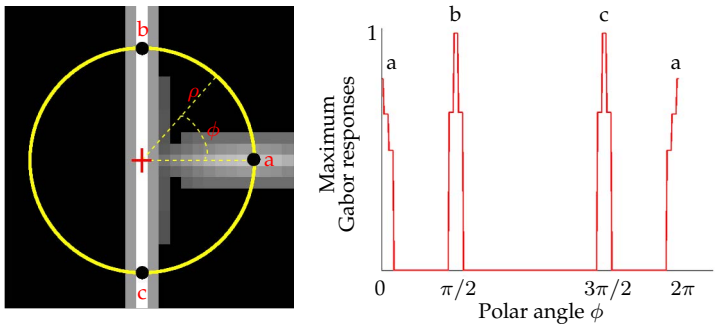


Figure 5.7: Configuration of a colour COSFIRE filter. (a) The gray-level of a pixel represents the maximum value superposition of the thresholded (at $t_1 = 0.4$) responses of a bank of symmetric Gabor filters (4 wavelenghts $\lambda \in \{3, 6, 10, 14\}$, 6 orientations $\theta = \left\{\frac{\pi i}{6}, i = 0...5\right\}$ and $\zeta = 0$) at that position. The red cross indicates the location of the point of interest (in this case selected by the user) and the yellow circle represents the locations considered around the point of interest for a given radius ρ, here $\rho = 10$. (b) Values of the maximum value superposition of the thresholded responses of the bank of Gabor filters along the concerned circle. The three local maxima in the plot are respectively labelled and marked with black dots in (a). The local positions of the local maxima in (a) relative to the centre of the filter (ρ_i, ϕ_i) and the wavelength and orientation (λ_i, θ_i) of the Gabor filter that produced such response describe partly a tuple.

kernel $K_{\lambda_i,\theta_i,\zeta}$ around the location (ρ_i, ϕ_i) and perform a pixel-wise multiplication of the kernel by a colour channel of the prototype image P_c. Then, we normalise the result. Thus, we obtain a colour description value for each colour channel at the considered location, γ_{c_i}.

$$\gamma_{c_i} = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n} P_c(x_i + k - 1, y_i + l - 1) K_{\lambda_i,\theta_i,\zeta}(k,l)}{\sum_{k=1}^{m}\sum_{l=1}^{n} K_{\lambda_i,\theta_i,\zeta}(k,l)} \tag{5.4}$$

where m and n are the rows and columns of the kernel $K_{\lambda_i,\theta_i,\zeta}$ respectively and (x_i, y_i) the Cartesian coordinates of (ρ_i, ϕ_i). We compute this average rather than directly using the value of the pixel (ρ_i, ϕ_i) at each colour channel to avoid that possible noisy values of pixels may deeply affect the colour description. Figure 5.8 shows the regions of the prototype pattern considered to compute the colour descriptions. For symmetric Gabor filters, both kernels $K_{\lambda_i,\theta_i,\zeta}$ for values of $\zeta \in \{0, \pi\}$ are identical, so anyone can be taken to describe the colour. Since we are using the central part of a symmetric Gabor filter of wavelength λ_i, we ensure that the colour description is computed at a region of width equals to, at most, $\lambda_i/2$, which is the width of the line that the method localises. For antisymmetric Gabor filters, we use the kernel $K_{\lambda_i,\theta_i,\zeta}$ with the value of $\zeta \in \{\pi/2, 3\pi/2\}$ in which the Euclidean dis-

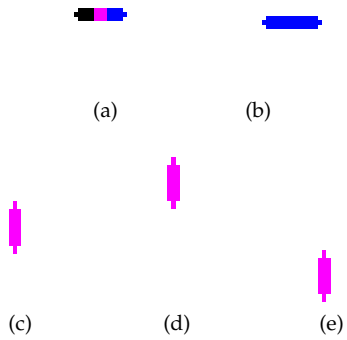


Figure 5.8: Regions of the prototype pattern, Fig. 5.6b, considered to compute the colour description in each contour part (white pixels are not considered). On the one hand, (a) and (c) correspond to the contour parts in the centre of the prototype. On the other hand, (b) corresponds to the labelled point 'a', (d) to 'b' and (e) to 'c' in Fig. 5.7a.

tance from the centroid of the kernel to the interest point is minimum when both kernels are centred around the location (ρ_i, ϕ_i). In this way, we describe the part of the prototype that is closer to the centre of the colour COSFIRE filter.

A set of seven parameter values or tuple $p_i = (\lambda_i, \theta_i, \rho_i, \phi_i, \gamma_{1_i}, \gamma_{2_i}, \gamma_{3_i})$ specifies the properties of a contour part. The set $S_f = \{p_i | i = 1 \dots n_c\} = \{(\lambda_i,\theta_i,\rho_i,\phi_i,\gamma_{1_i},\gamma_{2_i},$ $\gamma_{3_i}) \mid i = 1 \dots n_c\}$ denotes the parameter values combinations which fulfil the above conditions. The subscript f stands for the prototype pattern around the selected point of interest and n_c is the number of localised contour parts.

For the prototype shown in Fig. 5.6b and Fig. 5.7a, with two values of the parameter ρ ($\rho = \{0, 10\}$), this method results in five contour parts with parameter values specified by the tuples in the set shown in Table 5.3. The second tuple $(\lambda_2 = 10, \theta_2 = \pi/2, \rho_2 = 10, \phi_2 = 0, \gamma_{1_2} = 0, \gamma_{2_2} = 0, \gamma_{3_2} = 1)$ describes a contour part with a width of $(\lambda_2/2 =)$ 5 pixels and an orientation of $\theta_2 = \pi/2$ that can be detected by a Gabor filter with preferred wavelength $\lambda_2 = 10$ and orientation $\theta_2 = \pi/2$, at a position of $\rho_2 = 10$ pixels to the right $(\phi_2 = 0)$ of the point of interest and with RGB colour description $[\gamma_{1_2} = 0, \gamma_{2_2} = 0, \gamma_{3_2} = 1] = [0, 0, 1]$ which is pure blue. This location is marked by the label 'a' in Fig. 5.7a. This selection is the result of the presence of a horizontal blue line to the right of the centre of prototype that is used for the configuration of the filter. The structure of the colour COSFIRE filter is represented in Fig. 5.9.

Table 5.3: Set of tuples that describe the contour parts of the prototype in Fig. 5.6b and 5.7a.

$$
\begin{array}{llllllll}
S_f = \{ & & & & & & & \\
(\lambda_1 = 10, & \theta_1 = \pi/2, & \rho_1 = 0, & \phi_1 = 0, & \gamma_{1_1} = 0.2, & \gamma_{2_1} = 0, & \gamma_{3_1} = 0.6), \\
(\lambda_2 = 10, & \theta_2 = \pi/2, & \rho_2 = 10, & \phi_2 = 0, & \gamma_{1_2} = 0, & \gamma_{2_2} = 0, & \gamma_{3_2} = 1), \\
(\lambda_3 = 6, & \theta_3 = 0, & \rho_3 = 0, & \phi_3 = 0, & \gamma_{1_3} = 1, & \gamma_{2_3} = 0, & \gamma_{3_3} = 1), \\
(\lambda_4 = 6, & \theta_4 = 0, & \rho_4 = 10, & \phi_4 = \pi/2, & \gamma_{1_4} = 1, & \gamma_{2_4} = 0, & \gamma_{3_4} = 1), \\
(\lambda_5 = 6, & \theta_5 = 0, & \rho_5 = 10, & \phi_5 = 3\pi/2, & \gamma_{1_5} = 1, & \gamma_{2_5} = 0, & \gamma_{3_5} = 1) \\
\} & & & & & & &
\end{array}
$$

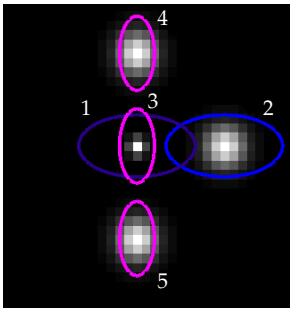


Figure 5.9: Structure of the colour COSFIRE filter for the prototype in Fig.5.6b. Each of the numbered ellipses represent a tuple of the set of contour parts shown in Table 5.3 labelled with the same identification numbers. The wavelengths and orientations of the Gabor filters at the local positions of the contour parts and the colours described for each contour part are taken into account for the representation. This filter is trained to detect the spatial local arrangement and colour of five contour parts. The bright blobs are intensity maps of the Gaussian functions that will be used in the application step for blurring the responses of the Gabor filters.

Application of a colour COSFIRE filter for vertex localisation

To obtain the response for line detection, we apply a bank of Gabor filters to an input image with the pairs of values (λ, θ) that form the tuples of the set S_f. To compute the responses for colour description, we apply Gaussian kernels to measure the similarity between the colour descriptions of the configuration and the ones from the input image. For each pixel, we obtain the responses at the local positions (ρ_i, ϕ_i) of S_f from the considered pixel in terms of lines detection and colour description. Since we want to achieve strong responses both for line detection and for color description for each contour part, we multiply the two responses. The output of the colour COSFIRE filter for each pixel in the image can be computed as a combination of all responses for the different contour parts defined in the configur-

ation step. The concerned responses for each contour part are in different positions (ρ_i, ϕ_i) with respect to the filter centre, thus we first shift them appropriately so that they come together to the filter centre. In the following, we explain in detail these steps.

Line/edge detection. We compute the responses of a bank of 2D Gabor filters applied to each colour channel of the input image for the pairs of values (λ_i, θ_i) of the set S_f and for both phase offset values ζ. If symmetric Gabor filters were used in the configuration selection, $\zeta = \{0, \pi\}$, otherwise if antisymmetric filters were applied, $\zeta = \{\pi/2, 3\pi/2\}$. Both values of ζ are analysed because we want the method to localise the pattern of interest independently of the background. In the same way as for the configuration of the colour COSFIRE filter, we apply two consecutive L-infinity norms, along the colour channels and along the phase offset values. Then we threshold the responses at a fraction t_1 of the maximum response, resulting in a Gabor response $|g_{\lambda_i,\theta_i}(x,y)|_{t_1}$ for each tuple p_i in the set S_f. We also obtain the Kernels $K_{\lambda_i,\theta_i,\zeta}$ associated with the Gabor filters of each tuple.

The Gabor filter responses are blurred to allow for some tolerance in the position of the contour parts. We define the blurring operation as a convolution, both along the rows and columns, of the thresholded Gabor responses $|g_{\lambda_i,\theta_i}(x,y)|_{t_1}$ with a rotationally symmetric Gaussian lowpass filter $G_\sigma(x,y)$ of size $1 \times n\sigma$ pixels with standard deviation σ. The standard deviation is a linear function of the distance ρ from the centre of the colour COSFIRE filter,

$$\sigma = \sigma_0 + \alpha\rho \tag{5.5}$$

where n, σ_0 and α are constants. The orientation bandwidth is broader with a higher value of α. The visual system of the brain inspired the choice of the linear function in Eq. 5.5, following Azzopardi and Petkov (2013c). The blurred response for a tuple p_i is

$$b_{\lambda_i,\theta_i,\rho_i}(x,y) = |g_{\lambda_i,\theta_i}(x,y)|_{t_1} * G_{\sigma_i}(x,y) * G'_{\sigma_i}(x,y) \tag{5.6}$$

Next, we shift the blurred responses of each tuple p_i by a distance of ρ_i in the opposite direction to ϕ_i. In polar coordinates, we can express this as $\rho_i\phi_i + \pi$, whereas in cartesian coordinates it is described as an increment $(\Delta x_i, \Delta y_i)$ where $\Delta x_i = -\rho_i \cos\phi_i$ and $\Delta y_i = -\rho_i \sin\phi_i$. We denote by $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ the blurred and shifted response of the Gabor filter specified by the tuple p_i in the set S_f:

$$s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) = b_{\lambda_i,\theta_i,\rho_i}(x - \Delta x_i, y - \Delta y_i) \tag{5.7}$$

Figure 5.10 shows the application of a colour COSFIRE filter to an input image

for line detection. The response of the colour COSFIRE filter for line detection is achieved by computing five blurred and shifted responses of two pairs of Gabor filters. Each pair of Gabor filters has the same parameters except for their phase offset values, which are complementary, and they are both combined by taking the maximum value of the response for every pixel. Each of the five responses corresponds to each contour part found in the configuration.

Colour description. First, for each tuple p_i, we convolve each colour channel of the input image I_c with the corresponding sliding kernels $K_{\lambda_i,\theta_i,\zeta}$ and then we normalise the results.

$$v_{\lambda_i,\theta_i,c}(x,y) = \frac{I_c(x,y) * K_{\lambda_i,\theta_i,\zeta}(x,y)}{\sum_{k=1}^{m}\sum_{l=1}^{n} K_{\lambda_i,\theta_i,\zeta}(k,l)} \tag{5.8}$$

where k and l are the rows and columns of the kernel $K_{\lambda_i,\theta_i,\zeta}$ respectively.

For symmetric Gabor filters, since the same kernel is computed for both values of ζ, any of them can be used as the sliding kernel. For antisymmetric Gabor filters, we compute the convolutions with $K_{\lambda_i,\theta_i,\zeta=\pi/2}$ and $K_{\lambda_i,\theta_i,\zeta=3\pi/2}$.

We denote by $h_{p_i}(x,y)$ the response for colour description of the tuple p_i in the set S_f. We compute $h_{p_i}(x,y)$ by applying a Gaussian kernel that measures the similarity between the colours of the prototype contour part and the colours of the input image in each colour channel as in Eq. 5.9.

$$h_{p_i}(x,y) = exp^{-\frac{\sum_{c=1}^{3}[v_{\lambda_i,\theta_i,c}(x,y)-\gamma_{c_i}]^2}{2\sigma_g^2}} \tag{5.9}$$

where σ_g is the standard deviation of the colour Gaussian kernel.

For antisymmetric Gabor filters, we compute a response for colour description $h_{p_i}(x,y)$ for each Gabor kernel and then we obtain the maximum value along both responses for each pair of corresponding pixels (x_j, y_j).

$$h'_{p_i}(x,y) = \max_{j,k}\{h_{p_i}(x_j,y_k)|\zeta = \pi/2,\ h_{p_i}(x_j,y_k)|\zeta = 3\pi/2\} \tag{5.10}$$

Afterwards, we blur the response for colour description.

$$h''_{p_i}(x,y) = h'_{p_i}(x,y) * G_{\sigma_i}(x,y) * G'_{\sigma_i}(x,y) \tag{5.11}$$

And finally, we shift the blurred response for colour description a distance of ρ_i in the opposite direction to ϕ_i.

$$\hat{h}_{p_i}(x,y) = h''_{p_i}(x - \Delta x_i, y - \Delta y_i) \tag{5.12}$$

Figure 5.11 shows the application of a colour COSFIRE filter to an input im-

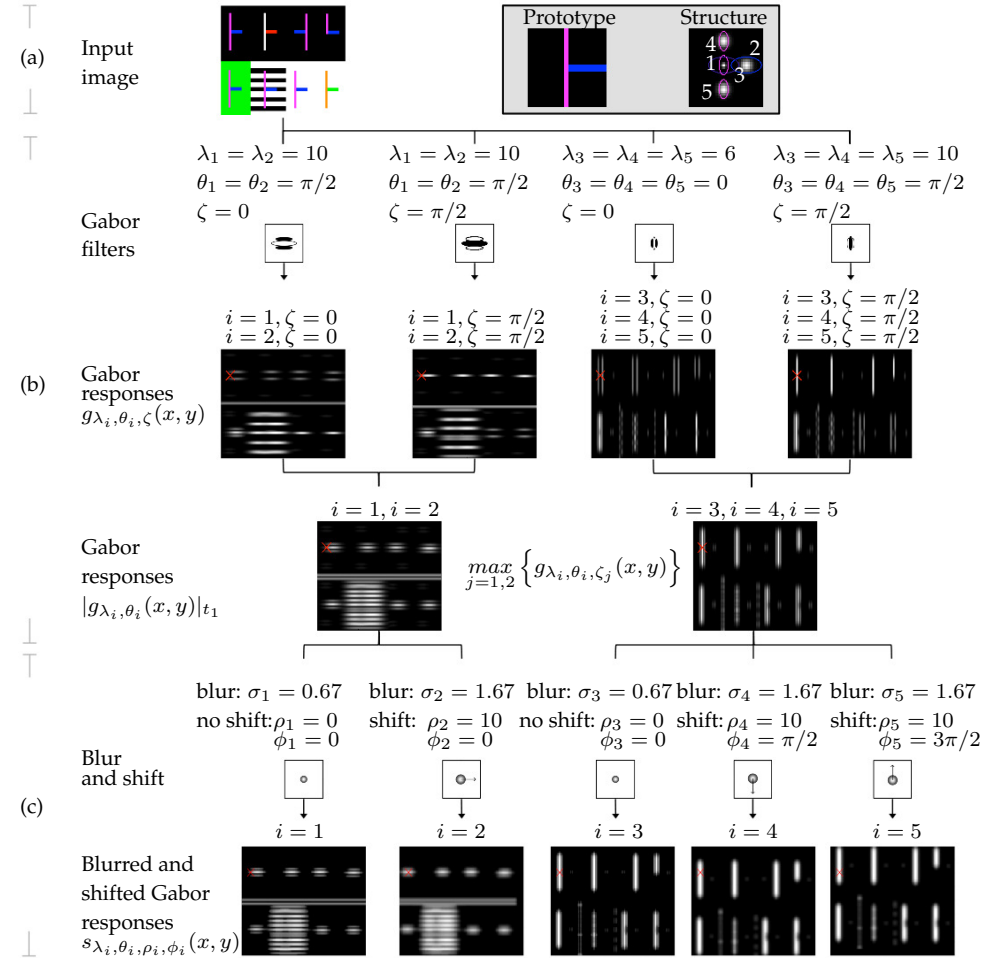


Figure 5.10: The '×' marker indicates the location of the point of interest. (a) Input image. The framed area shows (left) the enlarged pattern of interest selected for the configuration and (right) the structure of the colour COSFIRE filter that was configured for this pattern. (b) Each contour part of the prototype pattern is detected by the combination of the responses of a pair of symmetric Gabor filters with preferred values of wavelength λ_i and orientation θ_i and phase offsets $\zeta_i = \{0, \pi\}$. Two of the contour parts ($i = \{1, 2\}$) are detected by one pair of Gabor filters and the other three parts ($i = \{1, 2, 3\}$) are detected by another pair of Gabor filters. Thus, only two pairs of distinct Gabor filters are chosen from the filter bank. (c) The thresholded response $|g_{\lambda_i,\theta_i}(x,y)|_{t_1}$ (here $t_1 = 0.2$) is then blurred (here $n = 6$) and later shifted by $(\rho_i \phi_i + \pi)$ in polar coordinates.

age for colour description. The response of the colour COSFIRE filter for colour description is achieved by computing five blurred and shifted responses of three Gaussian kernel similarities. Each of the five responses corresponds to each contour part found in the configuration.

Response of a colour COSFIRE filter. We define the response of a colour COSFIRE filter $r_{S_f}(x,y)$ as the weighted geometric mean of the Hadamard product of the blurred and shifted thresholded Gabor filter responses, $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$, by the blurred and shifted Gaussian colour responses, $\hat{h}_{p_i}(x,y)$, that correspond to the properties of the contour parts described in S_f:

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \circ \hat{h}_{p_i}(x,y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_f|}\omega_i} \tag{5.13}$$

$$\omega_i = exp^{-\frac{\rho_i^2}{2\sigma'^2}} \tag{5.14}$$

$$\sigma' = \left(-\rho_{max}^2/2ln\tau\right)^{1/2} \tag{5.15}$$

$$\rho_{max} = \max_{i \in 1 \ldots |S_f|} \{\rho_i\} \tag{5.16}$$

where $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \circ \hat{h}_{p_i}(x,y)$ stands for the Hadamard product of $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ and $\hat{h}_{p_i}(x,y)$. When $1/\sigma' = 0$, the weighted geometric mean becomes a standard geometric mean and all the contour parts responses $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \circ \hat{h}_{p_i}(x,y)$ have the same contribution. On the contrary, for $1/\sigma' > 0$ the contribution of the contour parts decreases with an increasing value of the corresponding parameter ρ. In particular, we achieve a maximum value $\omega = 1$ of the weights in the centre ($\rho = 0$), and minimum value $\omega = \tau$ in the periphery ($\rho = \rho_{max}$).

Finally, we threshold the response of the colour COSFIRE filter at a fraction t_3 of its maximum across all image coordinates (x,y), $0 \leq t_3 \leq 1$.

$$r(x,y) = \left| r_{S_f}(x,y) \right|_{t_3} \tag{5.17}$$

Figure 5.12 shows the application of a colour COSFIRE filter to an input image for the localisation of a colour vertex. The output of the colour COSFIRE filter is the weighted geometric mean of the Hadamard multiplication of five blurred and shifted responses of two pairs of Gabor filters and five blurred and shifted responses of three convolutions. The filter responds at points where there is a pattern identical or similar to the prototype pattern (Fig. 5.6b) and at the point of interest of the proto-

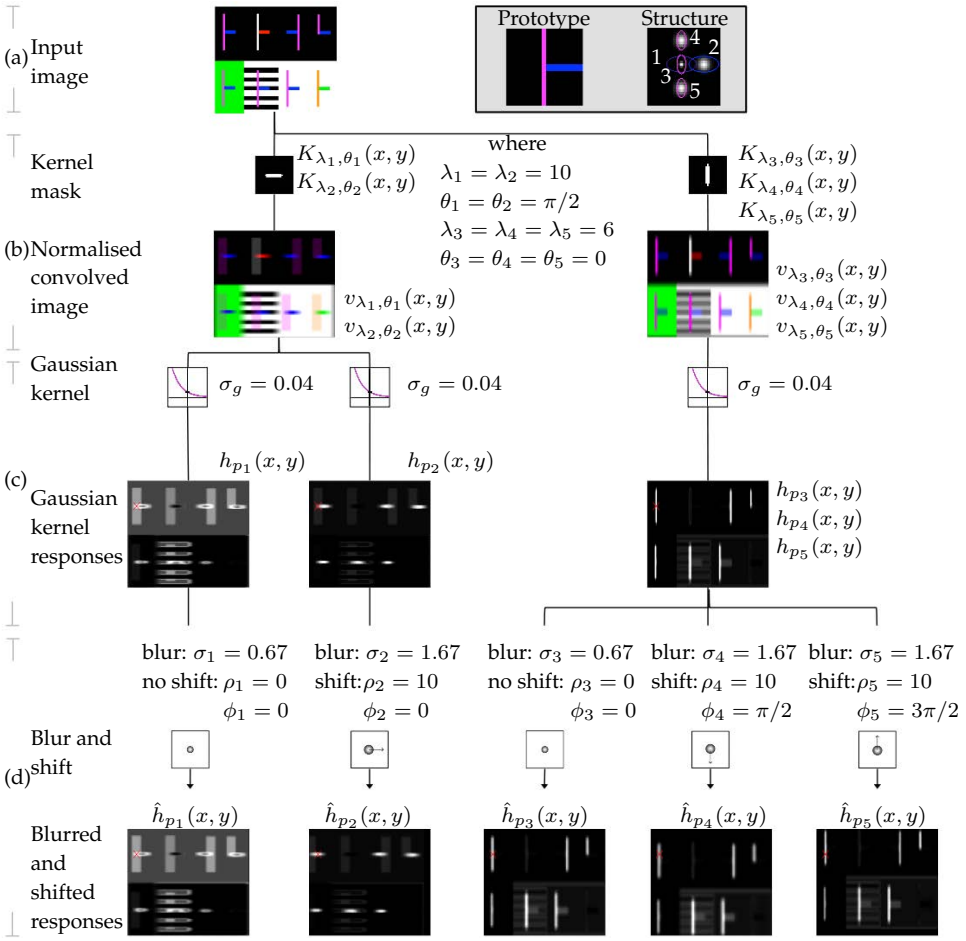


Figure 5.11: The '×' marker indicates the location of the point of interest. (a) Input image. The framed area shows (left) the enlarged pattern of interest selected for the configuration and (right) the structure of the colour COSFIRE filter that was configured for this pattern. (b) The image is convolved with two sliding kernels defined by the two Gabor filters used for line detection and then is normalised. (c) We compute the Gaussian kernel similarity between the colours of the prototype contour part and the colours of the input image. There are only three tuples with unique values of $(\lambda_i, \theta_i, \gamma_{1_i}, \gamma_{2_i}, \gamma_{3_i})$ and therefore only three similarities are obtained. (d) The Gaussian kernel responses are then blurred (here $n = 6$) and later shifted by $(\rho_i \phi_i + \pi)$ in polar coordinates.

type pattern despite the different colors and patterns of the background. Thus, we are getting strong responses in a given point to a local pattern that contains a hori-

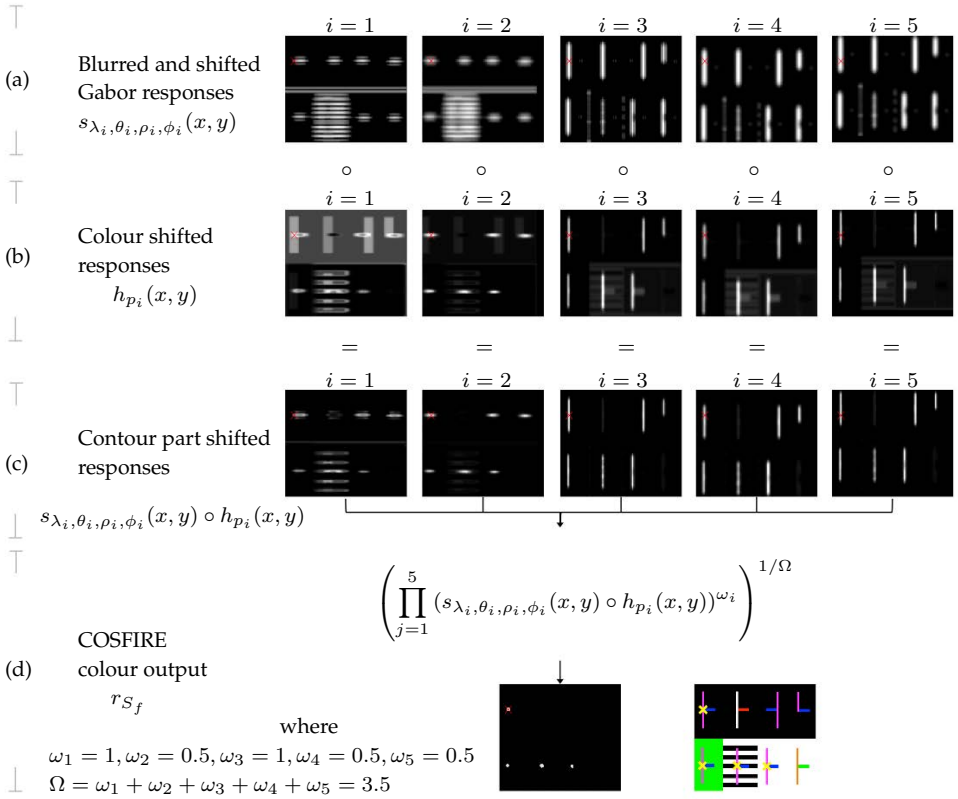
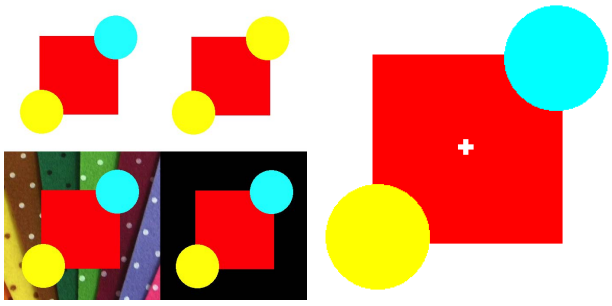


Figure 5.12: The '×' marker indicates the location of the point of interest. (a) Blurred and shifted Gabor responses for line detection as shown in Fig. 5.10(c). (b) Blurred and shifted Gaussian kernel responses for colour description as shown in Fig. 5.11. (c) We, then, multiply the blurred and shifted Gabor responses by the blurred and shifted colour responses achieving the final responses of each contour part. (f) Finally, the output of the colour COSFIRE filter is computed by the thresholded weighted geometric mean of all the contour part responses, here $\tau = 0.5$ and $t_3 = 0.7$. The five local maxima in the output of the colour COSFIRE filter correspond to the five similar vertices, in shape and colour, in the input image. They are found despite the different backgrounds.

zontal blue line to the right of the aforementioned point, a vertical pink line above and under the point and a horizontal bluish and a vertical pink lines at the point.



(a) (b)

Figure 5.13: 5.13a Synthetic input image of size 700×550 pixels. 5.13b Enlargement of the protytpe from which the colour COSFIRE filter will be obtained. It corresponds to the top left object in 5.13a. The white cross indicates the centre of the prototype, in this case automatically assigned as the centre of the ROI.

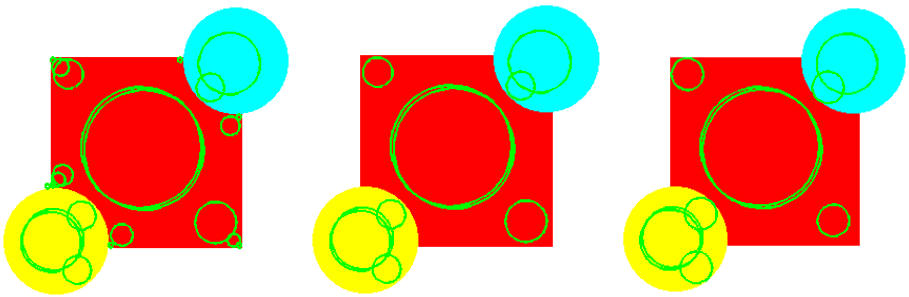
5.2.2. Method with application for colour object localisation

Overview

The previous method describes the colour of lines or edges but the colour of an object is also defined by the colour of its surface. We define a new set of tuples for the colour description of blobs in the surface of a prototypical object of interest. For each contour part of the new set of tuples, we compute the response of the colour description of blobs in an input image in the same way as the response for colour description of lines. The response of the colour COSFIRE filter is obtained by the Hadamard multiplication of the response for colour edge detection, explained in the previous section, and the weighted geometric mean of the blurred and shifted Gaussian kernel similarities for colour description of blobs.

Figure 5.13a shows an input image with four objects. We consider the top left object as the prototype of interest. The ROI that encompass the prototype is shown enlarged in Fig. 5.13b. The colour COSFIRE filter configured for this prototype will respond to the same and similar patterns in terms of shape and colours.

In the succeeding sections we explain the extraction of the new set of tuples for colour description of blobs, the extraction of the response of the colour description of blobs for an input image and the computation of the colour COSIFRE filter response.



(a) (b) (c)

Figure 5.14: The green circles show the keypoints found using SIFT detector for the prototype object of interest. (a) All unique SIFT keypoints detected. The radius of the circle represents the scale at which the keypoint was found. (b) Remaining keypoints after thresholding with $t_4 = 0.2$. (c) Keypoints with only three scales.

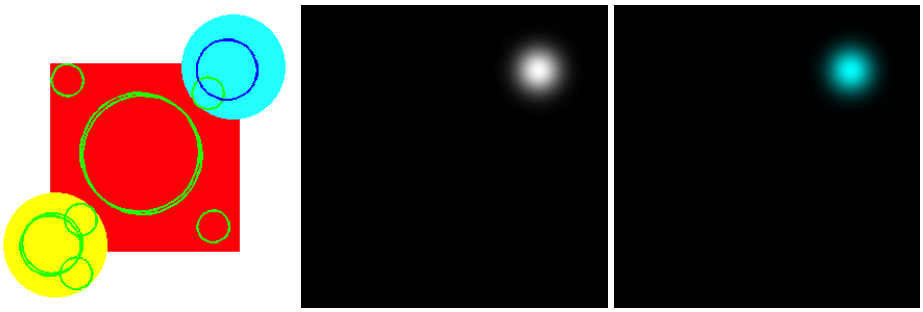
Configuration of a colour COSFIRE filter for object localisation

We use SIFT detector (Lowe, 2004) to look for stable keypoints in the prototype. SIFT is a blob detection method that localises regions in images that differ in properties compared to the surrounding regions. A SIFT keypoint is a circular image region described by a geometric frame of four parameters: the keypoint centre coordinates (x_j, y_j), its scale (that it is equal to the radius of the region) δ_j, and its orientation, (Vedaldi and Fulkerson, 2008). We are only interested in the coordinates and scale of the keypoints.

We apply SIFT detector to every channel of the input image I_c and consider the keypoints whose scale is greater than a fraction t_4 of the maximum scale across all keypoints. Then we cluster the remaining keypoints into three groups according to their scale values using k-means (Duda et al., 2000), and assign to each keypoint the mean scale value of the group to which they belong. This step is not essential but it allows to speed up later computations. Finally, only unique keypoints (δ_j, x_j, y_j) are kept, Fig. 5.14.

The point of interest of the prototype (x_p, y_p), which is the centre of the colour COSFIRE filter, can be either manually selected by the user or automatically assigned as the centre of the ROI. We compute the local polar coordinates (ρ_j, ϕ_j) of the keypoints (x_j, y_j) with respect to the point of interest of the prototype pattern.

$$(\rho_j, \phi_j) = \left(\sqrt{(x_j - x_p)^2, (y_j - y_p)^2}, \ \text{atan2}(y_j - y_p, x_j - x_p) \right) \tag{5.18}$$



(a) (b) (c)

Figure 5.15: Colour description of a blob. (a) Prototypical object of interest. SIFT keypoints are marked in green. We choose one keypoint, marked in blue, as example for the computation of the colour description of that blob. (b) Gaussian circular mask $K_{\delta_j,\rho_j,\phi_j}(x,y)$ for the keypoint marked in blue. (c) Result of the pixel-wise multiplication of the Gaussian circular mask by the prototypical object of interest. The colour description of this keypoint for RGB colour space results in $[\gamma_{c_1}=0,\gamma_{c_2}=1,\gamma_{c_3}=1]=[0,1,1]$ which is cyan colour.

where atan2 is the angle in radians between the positive x-axis of a plane and the point given by the coordinates (x_j,y_j) on it.

For each keypoint (δ_j,ρ_j,ϕ_j), we create a Gaussian circular mask $K_{\delta_j,\rho_j,\phi_j}(x,y)$ of radius δ_j centred at the corresponding locations (ρ_j,ϕ_j).

$$K_{\delta_j,\rho_j,\phi_j}(x,y)=\exp^{-\frac{x^2+y^2}{2(\delta/2)^2}} \tag{5.19}$$

We then perform a pixel-wise multiplication of the mask by each colour channel of the prototype P_c and then normalise the result. In this way, the pixels closer to the considered location have a stronger participation in the computation of the colour description of the blob than the ones at further distances, Fig. 5.15. Therefore, we obtain a colour description value for each colour channel γ_{c_j} at the considered keypoint (δ_j,ρ_j,ϕ_j).

$$\gamma_{c_j}=\frac{\sum_{k=1}^{m}\sum_{l=1}^{n}P_c(x_j+k-1,y_j+l-1)K_{\delta,x_j,y_j}(k,l)}{\sum_{k=1}^{m}\sum_{l=1}^{n}K_{\delta,x_j,y_j}(k,l)} \tag{5.20}$$

where m and n are the rows and columns of the kernel K_{δ,x_j,y_j} respectively and (xj,y_j) the Cartesian coordinates of (ρ_j,ϕ_j).

A set of six parameter values or tuple $p_j=(\delta_j,\rho_j,\phi_j,\gamma_{1_j},\gamma_{2_j},\gamma_{3_j})$ specifies the properties of a contour part in this new set $S'_f=\{p_j|j=1\dots n_k\}=\{(\delta_j,\rho_j,\phi_j,\gamma_{1_j},\gamma_{2_j},\gamma_{3_j})\mid j=1\dots n_k\}$. The subscript f stands for the prototype object of interest

Table 5.4: Three tuples that give examples of the colour description of blobs of the prototypical object of interest in Fig. 5.6b and 5.7a. A total of 12 tuples were automatically described.

S_f = {

($\delta_1 = 56.8$,	$\rho_1 = 0$,	$\phi_1 = 0$,	$\gamma_{1_1} = 1$,	$\gamma_{2_1} = 0$,	$\gamma_{3_1} = 0$),
($\delta_2 = 28.7$,	$\rho_2 = 120.9$,	$\phi_2 = -3\pi/4$,	$\gamma_{1_2} = 1$,	$\gamma_{2_2} = 1$,	$\gamma_{3_2} = 0$),
($\delta_3 = 28.7$,	$\rho_3 = 115.3$,	$\phi_3 = \pi/4$,	$\gamma_{1_3} = 0$,	$\gamma_{2_3} = 1$,	$\gamma_{3_3} = 1$),

}

around the point of interest and n_k is the number of detected keypoints.

We compute another set of tuples $S_f = \{p_i | i = 1 \dots n_c\} = \{(\lambda_i, \theta_i, \rho_i, \phi_i, \gamma_{1_i}, \gamma_{2_i}, \gamma_{3_i}) \mid i = 1 \dots n_c\}$ for the object of interest as in Section 5.2.1.2 using a bank of antisymmetric Gabor filters with $\lambda = 20$ and $\theta = \{0, \pi/6, \pi/3, \pi/2, 2\pi/3, 5pi/6\}$.

For the prototype shown in Fig. 5.13b, this method results in two sets of tuples. Regarding the colour edge description, we localise 67 contour parts in a set S_f. As for the colour description of blobs with $t_4 = 0.2$, we localise 12 contour parts or keypoints in a set S'_f. Table 5.4 indicates the parameter values for three of those 12 keypoins. The third tuple describes a keypoint with a scale of $\delta_3 = 28.7$ pixels that can be detected by a SIFT detector at position $\rho_3 = 120.9$ pixels to the top right-hand corner ($\phi_3 = \pi/4$) of the point of interest (centre of the ROI) and with RGB colour description $[\gamma_{1_3} = 0, \gamma_{2_3} = 1, \gamma_{3_3} = 1] = [0, 1, 1]$ which is cyan. This selection is the result of the presence of a cyan colour blob to the top right-hand corner of the centre of the prototype that is used for the configuration of the filter. This structure is represented in Fig. 5.16.

Application of a colour COSFIRE filter for object localisation

We obtain the response for colour description of blobs by applying Gaussian kernels to measure the similarity between the colour descriptions of blobs at the configuration and the ones of the input image. Thus, this computation shares the main steps with the one used for the colour evaluation of lines and edges. The output of the colour COSFIRE filter is computed as the Hadamard product of the output obtained in 5.2.1.3 for colour edge detection and the response for colour description of blobs.

Colour description of blobs. For each unique value of δ_j in the tuples of S'_f, we compute a Gaussian circular mask $K_{\delta_j}(x, y)$ that contains a circle of radius δ_j defined as in Eq. 5.19. Then we convolve each colour channel of the input image I_c with the mask $K_{\delta_j}(x, y)$ and normalise the results, as in Eq. 5.8.

We denote by $d_{p_j}(x, y)$ the response for colour description of blobs for the tuple p_j in the set S'_f. We compute $d_{p_j}(x, y)$ by applying a Gaussian kernel that measures

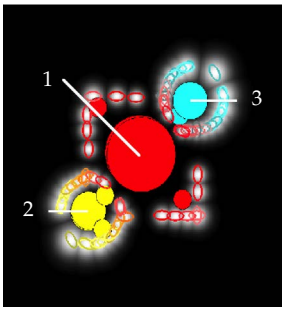


Figure 5.16: Structure of the colour COSFIRE filter for colour description of blobs of the prototypical object in Fig.5.13b. Each of the numbered circles represent a tuple of the set of contour parts shown in Table 5.4 with the same identification numbers. The wavelengths and orientations of the Gabor filters at the local positions of the contour parts and the colours described for each contour part are taken into account for the representation of the ellipses. The bright blobs are intensity maps of the Gaussian functions that are used in the application step for blurring the responses of the Gabor filters. The scale and colour described for each contour part for the colour description of blobs are considered for the representation of the circles. This filter is trained to detect the spatial local arrangement and colour of two sets of contour parts, one for colour edges and another for colour blobs.

the similarity between the colours of the contour part defined by the tuple p_j and the colours of the corresponding normalised and convolved input image along each colour channel, as in Eq. 5.9.

Afterwards, we blur the colour response, Eq. 5.11, and shift the blurred colour response a distance of ρ_j in the opposite direction to ϕ_j, Eq. 5.12, obtaining $\hat{d}_{p_j}$.

Response of a colour COSFIRE filter. We define the output $r_{S'_f}(x,y)$ of a colour COSFIRE filter for colour description of blobs in an object of interest as the weighted geometric mean of the blurred and shifted Gaussian similarity responses $d_{p_j}(x,y)$ that correspond to the properties of the contour parts described in S'_f:

$$r_{S'_f}(x,y) \stackrel{\text{def}}{=\!=} \left(\prod_{j=1}^{|S'_f|} \left(\hat{d}_{p_j}(x,y) \right)^{\omega_j} \right)^{1/\sum_{j=1}^{|S'_f|} \omega_j} \tag{5.21}$$

where ω_j is defined in Eq. 5.14.

We compute the response of the colour COSFIRE filter $r(x,y)$ as the thresholded Hadamard product of the responses for colour edge detection and for colour description of blobs:

$$r(x,y) \stackrel{\text{def}}{=\!=} \left| r_{S_f}(x,y) \circ r_{S'_f}(x,y) \right|_{t_5} \tag{5.22}$$

where $||_{t_5}$ stands for thresholding the response at a fraction t_5 of its maximum across all image coordinates (x, y).

Figure 5.17 shows the application of a colour COSFIRE filter for localisation of colour objects. The output of the colour COSFIRE filter is the Hadamard product of the weighted geometric mean of 12 responses for colour description of blobs and the weighted geometric mean of 67 responses for colour edges detection. The filter responds at points where there is an identical or similar pattern to the prototypical object of interest (Fig. 5.13b) and at the point of interest of the prototypical object of interest despite the different colors and patterns of the background. Thus, we are getting strong responses in a given point to a local pattern that contains a red square centred at the aforementioned point, a yellow circle at the bottom left-hand corner of the square and a cyan circle at the top right-hand corner of the square.

For the achievement of invariance to rotation, scale, reflection and contrast inversion of the colour COSFIRE filter, we refer the reader to (Azzopardi and Petkov, 2013c).

5.2.3. Experiments and results

dataset

We use COIL-100 public benchmark in our experiments. It consists of colour images of 100 object classes of size 128×128. 72 images of each object were taken which sums up to 7200 images for the whole dataset. The images were obtained by placing the objects on a turntable and taking a snapshot every 5°. The objects have a wide variety of complex geometric and pose characteristics. Images does not present occlusion, background clutter and illumination changes. Figure 5.18 shows the image taken at 0° for all objects of COIL-100 whereas Fig. 5.18 shows the viewpoints of three objects at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°.

Experimental set up and results

We configure one colour COSFIRE filter per object class for the image with rotation angle of 0°. We also configure standard COSFIRE filters for the same images. We use the same parameters for both colour and standard COSFIRE filters. We created a bank of Gabor filters with wave length $\lambda = 5$, orientations $\theta = \{0, \pi/8, ...\pi - \pi/8\}$, phase offsets $\zeta = \{\pi/2, 3\pi/2\}$ and aspect ratio 0.4. We set thresholds $t_1 = 0.1$, $t_2 = 0.75$ and $t_3 = 0$, and parameters related with the standard

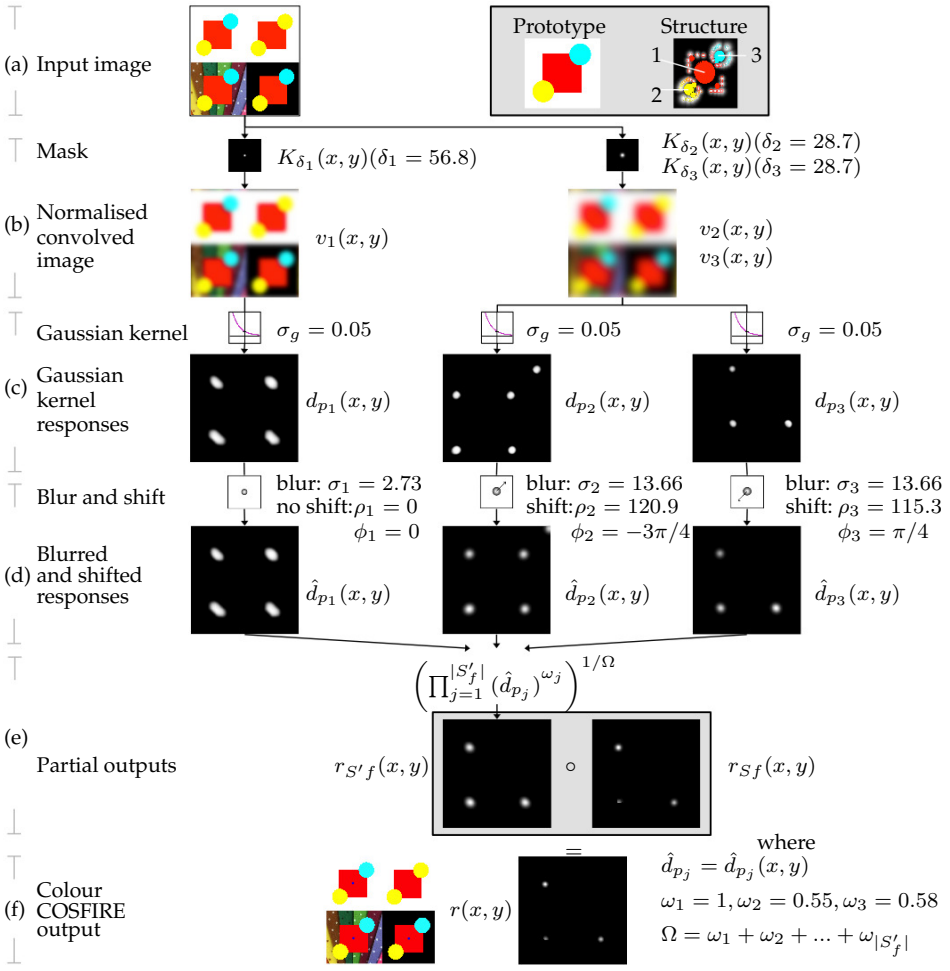


Figure 5.17: Application demonstration for localisation of colour objects. (a) Input image, prototype and structure of the colour COSFIRE filter. Numbers indicate three tuples in S'_f for which we illustrate this application. (b) Normalised convolution of the input image by a Gaussian circular mask of radius the scale of the contour part considered. (c) Similarity between the colours of the contour parts and the colours in the input image by a Gaussian kernel. (d) We blur and shift the previous responses. (e) The output of the colour COSFIRE filter for colour description of blobs is obtained as a weighted geometric mean of the blurred and shifted responses, $r_{S'_f}$. (f) We compute the output of the colour COSFIRE filter as the Hadamard product of $r_{S'_f}$ and the output for colour edges detection, r_{S_f}. The three local maxima in this output correspond to the three similar objects in the input image.

Figure 5.18: COIL dataset. Images taken at 0° of each object class. These are the objects considered for the configuration of colour COSFIRE filters.

deviation of the blurring function equal to $\sigma_0 = 0.83$ and $\alpha = 0.1$. We obtain the output of a COSFIRE filter by the geometric mean, thus $\omega = 1$. For colour description, we set $\sigma_g = 0.04$.

Figure 5.20 shows examples of the structures of the colour COSFIRE filters. The structures of the standard COSFIRE filters present the exact same tuples for contour description but without colour information.

We apply each configured COSFIRE filter to the whole dataset and compute pre-

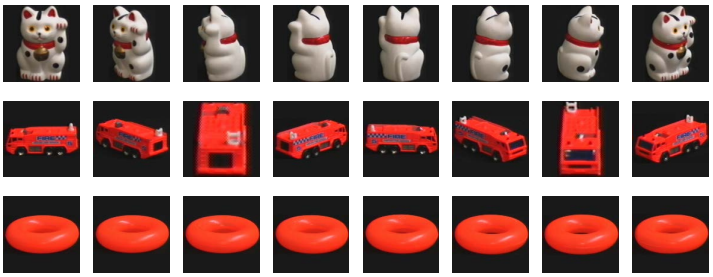


Figure 5.19: Viewpoints of three objects of COIL dataset at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315°.

cision and recall at every position in the retrieved hit list. We calculate the average precision, AveP, which is the area under a precision-recall curve, as

$$\text{AveP} = \frac{\sum_{k=1}^{n}(\text{P@k} \times \text{rel}[k])}{\text{number of relevant images}} \tag{5.23}$$

where k is the rank in the sequence of retrieved images, n is the number of retrieved images, P@k is the precision at cut k in the hit list and rel$[k]$ is 1 if the kth hit in the hit list is relevant and 0 otherwise.

Figure 5.21 shows plots of some precision-recall curves both for colour and standard COSFIRE filters. Table 5.5 indicates the average precisions obtained for each object class with both colour and standard COSFIRE filters. Colour COSFIRE filters have a higher distinctiveness power than standard COSFIRE filters since they always obtained higher average precisions.

We compute the mean average precision, MAP, for all the queries of the dataset as the mean of the average precision scores for each query,

$$\text{MAP} = \frac{\sum_{q=1}^{Q} \text{AveP}(q)}{Q} \tag{5.24}$$

where Q is the number of queries.

We also obtain the maximum harmonic mean of precision and recall for each query of the dataset. We compute the mean harmonic mean of precision and recall, MFScore, as the mean of the maxima harmonic means of precision and recall for all queries of the dataset. Mean precision, MPrecision, and mean recall, MRecall, are the means of the precisions and recalls, respectively, that obtained the maxima harmonic means for all queries of the dataset.

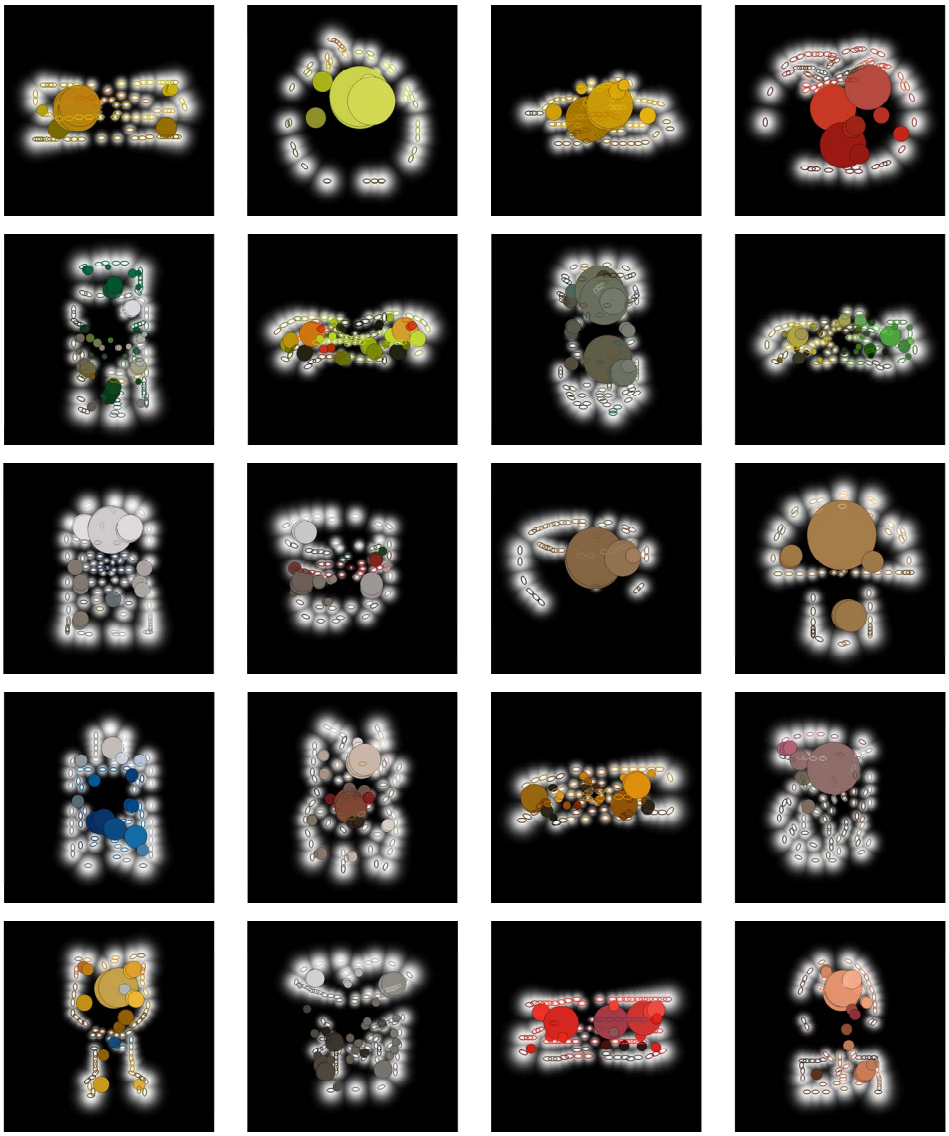


Figure 5.20: Structures of the colour COSFIRE filters for the first 20 classes of COIL dataset. The wavelengths and orientations of the Gabor filters at the local positions of the contour parts and the colours described for each contour part are taken into account for the representation of the ellipses. The bright blobs are intensity maps of the Gaussian functions that are used in the application step for blurring the responses of the Gabor filters. The scale and colour described for each contour part for the colour description of blobs are considered for the representation of the circles.

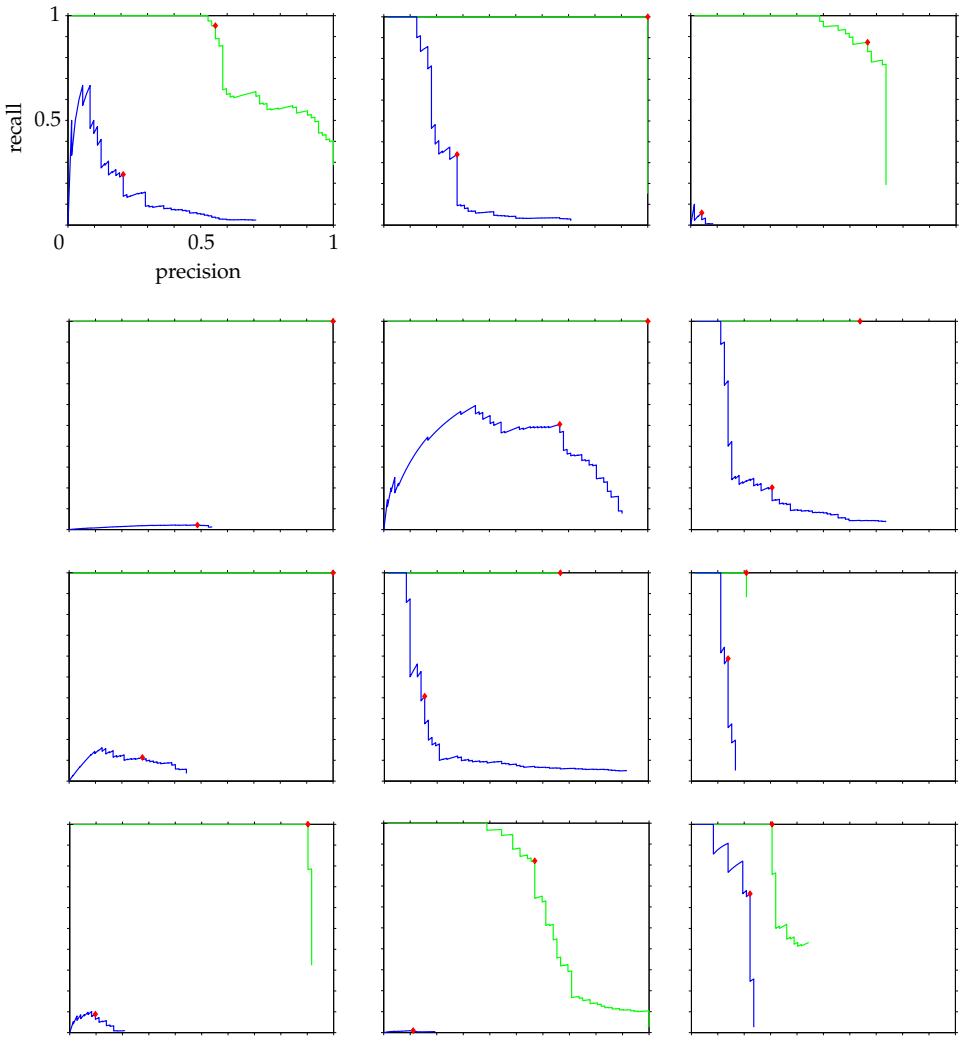


Figure 5.21: Precision-recall curves for the first 20 classes of COIL dataset. In green, precision-recall curves of the colour COSFIRE filters. In blue, precision-recall curves of the standard COSFIRE filters. Red diamonds indicate the maxima harmonic means of precision and recall.

Table 5.6 shows the values of MAP, MFScore, MPrecision and MRecall for CBIR demonstrating the effectiveness of colour COSFIRE filters with respect to standard COSFIRE filters.

We also evaluate the performance of COSFIRE filters as a classification problem.

Table 5.5: Average precisions of the 100 classes, Obj, of COIL dataset for colour COSFIRE filters, C (for colour), and COSFIRE filters, G (for gray).

Obj	C	G	Obj	C	G	Obj	C	G	Obj	C	G
1	0.81	0.13	**26**	0.92	0.16	**51**	0.47	0.27	**76**	1.00	0.10
2	1.00	0.23	**27**	0.93	0.03	**52**	0.64	0.16	**77**	0.20	0.09
3	0.71	0.00	**28**	0.44	0.00	**53**	1.00	0.02	**78**	0.77	0.01
4	1.00	0.01	**29**	0.90	0.35	**54**	0.50	0.12	**79**	0.41	0.14
5	1.00	0.38	**30**	1.00	0.00	**55**	0.30	0.17	**80**	0.15	0.09
6	0.64	0.21	**31**	0.47	0.26	**56**	1.00	0.01	**81**	0.86	0.57
7	1.00	0.05	**32**	0.96	0.24	**57**	0.38	0.09	**82**	0.61	0.01
8	0.67	0.19	**33**	0.57	0.17	**58**	1.00	0.01	**83**	1.00	0.04
9	0.21	0.13	**34**	1.00	0.06	**59**	1.00	0.20	**84**	1.00	0.07
10	0.91	0.01	**35**	1.00	0.19	**60**	0.30	0.04	**85**	0.45	0.11
11	0.66	0.00	**36**	0.71	0.00	**61**	0.44	0.05	**86**	1.00	0.01
12	0.37	0.20	**37**	0.94	0.56	**62**	0.29	0.05	**87**	1.00	0.37
13	0.26	0.12	**38**	0.43	0.00	**63**	0.44	0.00	**88**	0.96	0.13
14	0.28	0.14	**39**	0.38	0.16	**64**	0.60	0.45	**89**	0.68	0.10
15	0.80	0.13	**40**	0.72	0.09	**65**	0.24	0.13	**90**	0.56	0.02
16	0.82	0.44	**41**	0.55	0.01	**66**	0.73	0.17	**91**	0.63	0.06
17	0.61	0.08	**42**	0.97	0.02	**67**	0.75	0.19	**92**	0.58	0.19
18	0.24	0.03	**43**	0.99	0.01	**68**	0.71	0.18	**93**	0.54	0.13
19	0.47	0.06	**44**	0.35	0.03	**69**	0.86	0.10	**94**	1.00	0.02
20	0.97	0.01	**45**	0.26	0.01	**70**	0.90	0.16	**95**	1.00	0.11
21	0.50	0.07	**46**	0.53	0.05	**71**	1.00	0.08	**96**	0.88	0.18
22	0.26	0.19	**47**	1.00	0.21	**72**	1.00	0.42	**97**	0.24	0.14
23	0.90	0.01	**48**	0.67	0.04	**73**	1.00	0.20	**98**	1.00	0.06
24	1.00	0.06	**49**	0.92	0.11	**74**	0.35	0.14	**99**	0.94	0.24
25	1.00	0.92	**50**	1.00	0.09	**75**	0.20	0.06	**100**	0.40	0.09

Table 5.6: Mean average precision, MAP; mean harmonic mean, MFScore; mean precision, MPrecision; and mean recall, MRecall, of COIL dataset for colour COSFIRE filters, C, and COSFIRE filters, G.

	C	G
MAP	0.6970	0.1322
MFScore	0.7617	0.2241
MPrecision	0.9388	0.3217
MRecall	0.6822	0.3162

The responses of a given COSFIRE filter are divided by the maximum response obtained with that filter. A given image is classified to the class by which the COSFIRE filter that achieves the maximum response was configured. We compute a confusion

matrix where the value at location (i, j) is the number of images of class i classified as class j. Figures 5.22 and 5.23 show the confusion matrices of the colour COSFIRE filters and standard COSFIRE filters, respectively. The confusion matrix of the colour COSFIRE filters is less sparse than the one of the standard method, with high values at the diagonal and low values at the off-diagonal. The proposed colour-based approach yields 67.57% accuracy while the standard one achieves 21.69% accuracy, computing accuracy as the trace of the confusion matrix divided by the total number of images of the dataset.

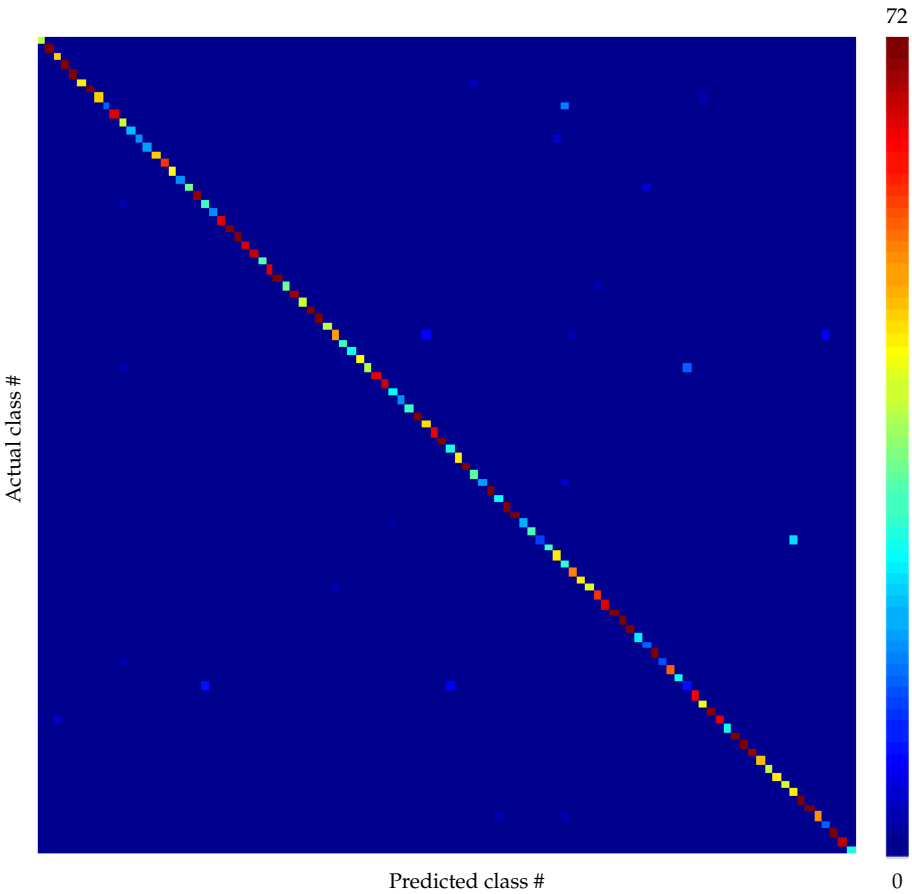


Figure 5.22: Confusion matrix for the proposed colour COSFIRE filters. The matrix is of size 100×100. The columns represent the instances in a predicted class and the rows the instances in an actual class.

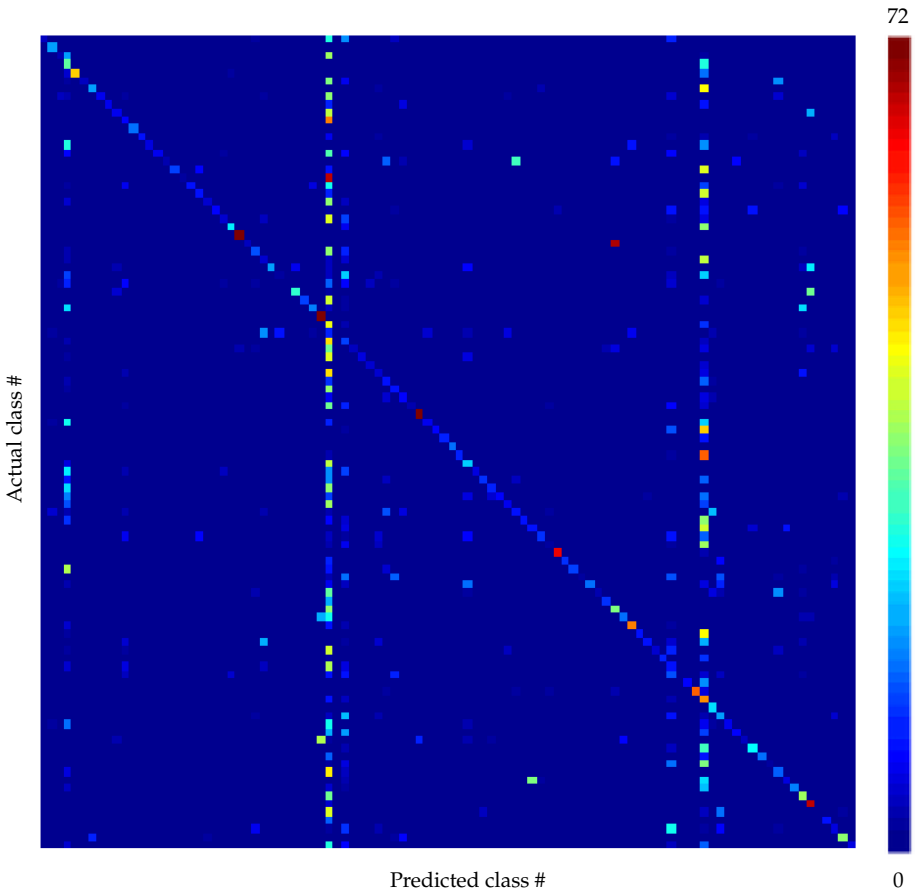


Figure 5.23: Confusion matrix for the standard COSFIRE filters. The matrix is of size 100×100. The columns represent the instances in a predicted class and the rows the instances in an actual class.

5.3. Conclusions

The contributions of the work presented in this chapter are two-fold. First, we evaluated different clustering configurations of SIFT keypoints in relation with their pose parameters: coordinates location, scale and orientation. On the one hand, we used the similarity measure of the closest pairs of keypoint descriptors. On the other hand, we used a Hough transform, with different parametrization values, to identify clusters of at least three points voting for the same pose of an object and we verified

the consistency of the pose parameters with the least squares algorithm. Second, we proposed colour COSFIRE filters that add colour description and discrimination to COSFIRE filters as well as providing invariance to background intensity. We presented colour COSFIRE filters both for patterns made up of colour lines and for patterns that are colour objects. Colour COSFIRE filters demonstrated to obtain higher retrieval and classification performance than the standard COSFIRE filters on COIL dataset.

All in all, in this section we contemplated two important tasks for object retrieval such as object matching and object localisation.

Chapter 6

Conclusions and outlook

6.1. Work summary

Three main applications have guided the work presented in this dissertation: the classification of boar spermatozoa according to the integrity of their acrosome heads; the localisation of cutting tools and identification of broken inserts in edge profile milling tools; and the retrieval of images containing certain objects for the *Advisory System Against Sexual Exploitation of Children (ASASEC)* project. Object recognition and image classification techniques, which have had a huge activity in the last years in the computer vision field, are required to provide a solution for these applications. In particular, in this thesis we have focused on the proposal of appropriate object recognition methods and retrieval techniques in these real applications.

The proportion of damage acrosomes in semen samples is usually estimated manually. Veterinary experts stain sperm samples and count the number of intact and damage acrosomes using a fluorescence microscope. Thus, the current process faces many drawbacks such as human mistakes or the requirement of expensive equipment. In this work we have analysed the integrity of boar acrosome spermatozoa describing their heads using invariant local features for the first time, as opposed to previous works that relied on global texture description.

Broken cutting tools may go on working without being detected and can cause a breakage of the head tool or even the milling machine itself. Tecoi utilises milling tools that contain a high number of inserts and that work under very aggressive conditions. Therefore, the identification of broken inserts is critical in this industrial process. We have proposed a method for the localisation of inserts and the identification of broken ones in such edge profile milling machines based on the specific geometry of these tools. Moreover, we have also presented a more general method for the localisation of inserts that can be automatically configured regardless of the appearance of the cutting tools and milling head tool.

In the ASASEC project, the retrieval of images and videos where some specific objects are present is one of the most challenging and important task to help fighting against sexual child exploitation. On the one hand, we have evaluated different

clustering configurations of SIFT keypoints for object matching in relation with their pose parameters: coordinates location, scale and orientation. On the other hand, we have presented a trainable keypoint detection operator, called colour COSFIRE filter, that firstly adds colour description and discrimination power to COSFIRE filters and, secondly, it provides invariance to background intensity.

In the rest of the chapter, the main conclusions of this work and future work lines are presented.

6.2. General conclusions

This dissertation has provided solutions to real applications using object recognition and image classification techniques.

Some specific conclusions that can be extracted from this work are:

1. *Invariant local features have been successfully applied, for the first time, in the assessment of sperm acrosome integrity.* We have demonstrated the success of applying SURF for the evaluation of the state of boar acrosomes as intact or damaged. SURF has obtained an averaged accuracy of 94.88%, 92.89% for the intact class and 96.86% for the damage one. It has been achieved classifying with k-NN algorithm, outperforming global texture descriptors and any other work presented to the date this results were published as a conference paper. Moreover, it has been observed that SURF and SIFT achieved higher accuracy rates for the damage class whereas global texture descriptors generally obtained better results for the intact one. Thus, a combination of both types of descriptions could improve the results obtained separately.

2. *In the same line of work, we have proposed an approach to classify SURF features, which produce several descriptors per image, with traditional SVM classifiers and without the use of BoW.* Classification of heads, using all their descriptors, has outperformed the assessment of single keypoint descriptors yielding an accuracy of 90.91% (94.94% and 86.87% for the intact and damage classes respectively). For the classification of heads, the intact class has obtained a higher accuracy than the damage one and the opposite situation has been yielded for the classification of points. We can conclude that keypoint descriptors detected in the damaged parts of the acrosome are more distinctive than the ones coming from the intact parts. However, damaged spermatozoa contain acrosome areas where the damage is not noticeable that could lead to keypoint mismatches. This approach can be easily extended to other ILF methods and classifier algorithms.

3. *The results of the proposed early fusion of ILF with global texture descriptors for the classification of the integrity of the acrosomes has outperformed the individual methods.* The concatenation of SURF with Legendre descriptors has achieved an accuracy of 95.56% (93.63% and 97.48% for the intact and damage classes respectively) when classifying with k-NN.

4. *A highly effective and efficient method for the localisation of cutting edges in milling machines has been presented.* Its output is a set of regions surrounding cutting edges, which can be used as input to other methods that perform quality assessment of the edges. It is based on applying a circular Hough transform to find the screws that fasten the inserts and edge detection and standard Hough transform to localise the cutting edge. It has obtained an accuracy of 99.61%, defining accuracy as the average of the fractions of the ground truth segments that lie inside ROIs of 20 pixels wide in images of 1280×960 pixels.

5. *A novel method has been introduced to describe and classify inserts as broken or unbroken with respect to the state of their cutting edges.* It computes the gradient magnitudes around the cutting edges and the deviation between the ideal and the real edges. The time required by this method for the inspection of the head tool is below the resting time of the machine. We have obtained an average harmonic mean $0.9143(\pm 0.079)$ with a precision $0.9661(\pm 0.073)$ and a recall $0.8821(\pm 0.134)$ in a publicly available dataset with 180 inserts by taking average results in 20 random validation sets.

6. *Another domain knowledge independent and more versatile method for the localisation of inserts has been presented.* It is more general than the previous one since it considers independently each image of the dataset. It is based on COSFIRE filters and it can be automatically configured regardless of the appearance of the inserts. A new metric, soft geometric mean, for the computation of the response of the COSFIRE filter has been introduced, outperforming the results obtained with the previous ones. This metric is based on geometric mean but it adds a small value to all entries and thus it provides tolerance to non-found contour parts. It has obtained a harmonic mean of 89.89%, with precision 92.39% and recall 87.52%, improving results of preceding works based on template matching.

7. *We have evaluated different clustering configurations of SIFT keypoints in relation with their pose parameters: coordinates location, scale and orientation.* On the one hand we have used the similarity measure of the closest pairs of keypoint descriptors. On the other hand we have used a Hough transform, with different parametrization values, to identify clusters of at least three points voting

for the same pose of an object and we verified the consistency of the pose parameters with the least squares algorithm. Higher precisions have been obtained without clustering at small cuts of the hit list whereas better precisions have been yielded with Lowe's clustering at high cuts. Results have been computed for a dataset of 614 images illustrating possible sceneries of ASASEC data base.

8. *Colour COSFIRE filters have been proposed.* They add colour description and discrimination power to COSFIRE filters as well as provide invariance to background intensity. Colour COSFIRE filters have been presented both for patterns made up of colour lines and for patterns that are colour objects. Colour COSFIRE filters have outperformed results for CBIR and classification tasks on COIL dataset with respect to standard COSFIRE filters.

6.3. Outlook

In this section, we summarise the main research lines that remain open for each studied application.

First, we discuss the classification of boar spermatozoa according to the acrosome integrity. The last work in the topic up to our knowledge, which I co-author, has been presented in (García-Olalla et al., 2015). It combines local and global texture descriptors and contour descriptors. Global texture description was obtained from the GLCM of the original image and the four sub-images of first level of decomposition with the DWT based on Haar wavelets. LBP and Fourier shape descriptors provided the local texture and the contour descriptions, respectively. An early fusion by concatenation of the descriptors was performed and the 10-fold classification using SVM backed by a least squares training algorithm and a linear kernel yielded an accuracy of 99.19% (harmonic mean equals to 99.12% of precision 99.42% and recall 98.84%). The solution for this application has already been delivered to the requesting company, Microptic.

Secondly, several lines of work are open for the development of a tool wear monitoring (TWM) system in edge profile milling machines:

1. The method for the localisation and identification of broken inserts that relies on the information of the several snapshots that capture the same insert under different poses can be improved by increasing the number of inserts localised in each individual image. This could be achieved by using, for example, a modification of the Hough transform for finding ellipses. The screws have a circular shape when the insert is placed in the centre of the image, but as the insert is placed on a side of the image the screw is seen as an ellipse. Another

option is to perform the localisation of the inserts with COSFIRE filters and then the localisation of cutting edges by edge detection and standard Hough transform. Having more views of the same insert could help to improve results since in some views there is low contrast of the inserts with respect to the background.

2. The results could be also improved with a better illumination system. Our method relies on edge detection and therefore achieving a good contrast is highly important. One possible solution would be to follow Pfeifer and Wiegers (2000) who captured the same insert under several lighting positions to combine contour information of all images. Otherwise, different settings of the illumination and capturing system could be evaluated.

3. Even though the breakage of the inserts is the most critical aspect to evaluate, it would be interesting that the TWM system also assesses the wear level of the inserts. In this way, we could predict possible future breakages of inserts or decide to change the inserts when they reach a high level of wear.

Finally, ASASEC project has been already delivered with satisfactory results. We put the effort in the appropriate programming of the product by following an object-oriented analysis and design making use of design patterns. Hence, future decisions of changing specifications will affect the programme in a limited way. However, the approach followed in the computer vision field was quite straightforward, with a more restricted development of new vision techniques. Colour COSFIRE filters can be exploited for many other applications related with object retrieval and object recognition in colour images or videos. The main drawback of colour COSFIRE filters is that they require a quite high number of convolutions, which depend on the application at hand, that are time consuming. Nevertheless, the implementation of the filters can be done in a parallel or distributed mode since most of the computations are independent from each other. Moreover, COSFIRE approach is not limited to the use of Gabor filter responses. In future, we will study object recognition using a combination of colour SIFT (Van de Sande et al., 2010) responses instead of Gabor filters responses. Previous tests using COSFIRE with SIFT for gray scale images are very promising. This approach can be very interesting for objects that contain distinctive blobs but less appropriate for objects in which contours are determinant.

Capítulo 7

Conclusiones y perspectiva

7.1. Resumen del trabajo

Tres aplicaciones han guiado el trabajo presentado en esta tesis: la clasificación de espermatozoides de verraco en función de la integridad del acrosoma de sus cabezas; la localización de plaquitas y la identificación de plaquitas rotas en fresadoras de corte de bordes; y la recuperación de imágenes que contienen ciertos objetos dados para el proyecto *Advisory System Against Sexual Exploitation of Children (ASASEC)*. Se requiere de las técnicas de reconocimiento de objetos y clasificación de imágenes, que han gozado de gran actividad durante los últimos años en el campo de visión computacional, para proveer una solución a estas aplicaciones. En concreto, en esta tesis nos hemos centrado en la propuesta de métodos adecuados de reconocimiento de objetos y técnicas de recuperación de imágenes para estas aplicaciones reales.

La proporción de acrosomas dañados en muestras de semen es generalmente estimada de manera manual. Los expertos veterinarios tiñen muestras de semen y cuentan el número de acrosomas intactos y dañados utilizando un microscopio de fluorescencia. Por tanto, el proceso actual conlleva muchos inconvenientes, tales como errores humanos o el requisito de un equipo de alto coste. En este trabajo hemos analizado la integridad de los acrosomas de los espermatozoides de verraco mediante la descripción de sus cabezas, utilizando características locales invariantes por primera vez, a diferencia de trabajos previos que se basaban en descripción de textura global.

Cabezales con plaquitas rotas pueden seguir trabajando sin que se detecte la existencia de dichas plaquitas defectuosas pudiendo causar daños en el cabezal o incluso en la fresadora. La empresa TECOI utiliza fresadoras que contienen un alto número de plaquitas y que trabajan bajo condiciones muy agresivas. Como consecuencia de ello, la identificación de plaquitas rotas tiene gran importancia en este proceso industrial. Hemos propuesto un método para la localización de plaquitas y la identificación de las que estén rotas en tales fresadoras de bordes basado en la geometría específica de estas herramientas. Además, hemos presentado un método

más general para la localización de plaquitas que puede ser configurado automáticamente independientemente de la apariencia de las plaquitas y el cabezal.

En el proyecto ASASEC, la recuperación de imágenes y vídeos en los que están presentes algunos objetos específicos es una de las tareas más difíciles e importantes para la lucha contra la explotación sexual de niños. Por un lado, hemos evaluado diferentes configuraciones de agrupamientos de los puntos clave SIFT para realizar la correspondencia de objetos en relación con los parámetros de la pose de dichos puntos clave: coordenadas de localización, escala y orientación. Por otro lado, hemos presentado un operador entrenable para la detección de puntos clave, llamado filtro color COSFIRE, que, en primer lugar, añade descripción de color y poder de discriminación a los filtros COSFIRE y, en segundo lugar, provee de invarianza a la intensidad del fondo de la imagen.

En el resto del capítulo, presentamos las principales conclusiones de este trabajo y las posibles líneas futuras de investigación.

7.2. Conclusiones generales

Esta tesis ha aportado soluciones a aplicaciones reales utilizando técnicas de reconocimiento de objetos y clasificación de imágenes.

Algunas conclusiones especificas que pueden extraerse de este trabajo son:

1. *Se han aplicado con éxito características locales invariantes, por primera vez, para evaluar de la integridad del acrosoma de espermatozoides.* Hemos demostrado la efeciencia de aplicar SURF para la clasificación del estado de los acrosomas de verraco como intactos o dañados. SURF obtuvo una tasa de acierto media del 94.88 %, 92.89 % para la clase intacta y 96.86 % para la clase dañada. Se obtuvo clasificando con el algoritmo k-NN, mejorando a los descriptores de textura global y a cualquier otro trabajo presentado en la fecha en la que estos resultados fueron publicados en forma de un artículo de conferencia. Además, se observó que SURF y SIFT obtuvieron mayores tasas de acierto para la clase dañada mientras que los descriptores de textura global generalmente consiguieron mejores resultados para la clase intacta. Por tanto, una combinación de ambos tipos de descriptores podría mejorar los resultados obtenidos de forma separada.

2. *En la misma línea de trabajo, hemos propuesto un método para la clasificación de características SURF, que utiliza varios descriptores por imagen, con clasificadores tradicionales SVM y sin el uso de BoW.* La clasificación de cabezas, utilizando todos sus descriptores, ha conseguido mejores resultados que la evaluación

de simples descriptores de puntos clave, obteniendo una tasa de acierto del 90.91 % (94.94 % y 86.87 % para las clases intacta y dañada respectivamente). En la clasificación de cabezas, la clase intacta ha obtenido una tasa de acierto más elevada que la dañada y la situación contraria se ha conseguido en la clasificación de puntos clave. Podemos concluir que los descriptores de puntos clave detectados en las partes dañadas del acrosoma son más distintivos que los que provienen de partes intactas. Sin embargo, los espermatozoides dañados contienen zonas donde el daño no es apreciable que pueden conducir a clasificaciones erróneas de puntos clave. Este enfoque puede ser extendido fácilmente para otros métodos ILF y algoritmos de clasificación.

3. *Los resultados de la fusión temprana de ILF con descriptores de textura globales propuesta para la clasificación de la integridad de los acrosomas ha mejorado a los métodos individuales.* La concatenación de los descriptores SURF y Legendre ha conseguido una tasa de acierto del 95.56 % (93.63 % y 97.48 % para las clases intacta y dañada respectivamente) cuando se clasifica mediante k-NN.

4. *Se ha presentado un método muy efectivo y eficiente para la localización de filos de corte en fresadoras.* Su salida es un conjunto de regiones alrededor de los filos de corte, las cuales pueden ser utilizadas como entrada para otros métodos que realicen una evaluación de la calidad de los filos. Está basado en la aplicación de una transformada circular Hough para encontrar tornillos que sujetan las plaquitas y en la detección de bordes y una transformada estándar Hough para la localización de los filos de corte. Ha obtenido una tasa de acierto del 99.61 %, definiendo tasa de acierto como la media de las fracciones de los segmentos de filo de corte reales que se encuentran dentro de ROIs de 20 píxeles de ancho en imágenes de 1280×960 píxeles.

5. *Se ha introducido un método novedoso para describir y clasificar plaquitas como rotas o no rotas en función del estado de sus filos de corte.* Calcula las magnitudes gradiente alrededor de los filos de corte y las desviaciones entre los filos reales e ideales. El tiempo requerido por este método para la inspección de un cabezal se encuentra por debajo del tiempo de reposo de la máquina. Hemos obtenido una media armónica de $0{,}9143(\pm0{,}079)$ con una precisión de $0{,}9661(\pm0{,}073)$ y una exhaustividad de $0{,}8821(\pm0{,}134)$ en un conjunto de imágenes que hemos dejado disponible públicamente de 180 plaquitas al calcular los resultados medios en 20 conjuntos de validación aleatorios.

6. *Se ha presentado otro método más versátil para la localización de plaquitas que además no requiere un conocimiento previo del dominio.* Es más general que el método anterior ya que considera de manera independiente cada imagen del cabezal.

Está basado en filtros COSFIRE y puede ser automáticamente configurado independientemente de la apariencia de las plaquitas. Se ha introducido una nueva métrica, media geométrica suave, para el cálculo de la respuesta del filtro COSFIRE, mejorando los resultados obtenidos con métricas previas. Esta función se basa en la media geométrica pero suma un pequeño valor a todas las entradas y, por consiguiente, aumenta la tolerancia a partes del contorno no encontradas. Ha obtenido una media armónica de 89.89 %, con una precisión de 92.39 % y una exhaustividad de 87.52 %, mejorando los resultados previos basados en correspondencia de plantillas.

7. *Hemos evaluado diferentes configuraciones de agrupamiento de puntos clave SIFT en función de sus parámetros de pose: coordenadas de localización, escala y roatación.* Por un lado, hemos utilizado la medida de similitud del par de descriptores más cercano. Por otro lado, hemos utilizado una transformada Hough, con diferentes parámetros, para identificar conjuntos de al menos tres puntos que voten por la misma pose de un objeto y hemos verificado la consistencia de dichos parámetros con un algoritmo de mínimos cuadrados. Se han obtenido precisiones más altas sin agrupamiento para cortes bajos de la lista de imágenes recuperadas, mientras que se han conseguido mejores precisiones con el agrupamiento de Lowe para cortes altos de dicha lista. Los resultados han sido calculados en un conjunto de 614 imágenes que ilustran un posible escenario de las colecciones de imágenes con las que se ha trabajado en ASASEC.

8. *Se han propuesto los filtros de color COSFIRE.* Añaden descripción de color y poder de discriminación a los filtros COSFIRE y, además, aportan invarianza a la intensidad del fondo de la imagen. Se han presentado los filtros de color COSFIRE para patrones compuestos por líneas de color y para patrones que son objetos de color. Los resultado de los filtros de color COSFIRE han mejorado los resultados para tareas de recuperación de imágenes basada en contenido y clasificación en el conjunto de imágenes de COIL con respecto a los filtros COSFIRE tradicionales.

7.3. Perspectiva

En esta sección, resumimos las principales líneas de trabajo que permanecen abiertas para cada aplicación estudiada.

En primer lugar, discutimos la clasificación de espermatozoides de verraco en función de la integridad del acrosoma. Hasta donde sabemos, el último trabajo en este ámbito y del cual soy coautora, ha sido presentado en (García-Olalla et al., 2015).

Combina descriptores de textura locales y globales y descriptores de contorno. La descripción de textura global se obtuvo a partir de la GLCM de la imagen original y de las cuatro sub-imágenes del primer nivel de descomposición con la DWT basada en wavelets de Haar. LBP y descriptores de forma de Fourier aportaron las descripciones de textura local y de contorno respectivamente. Se realizó una fusión temprana mediante concatenación de los descriptores y una clasificación 10-fold utilizando SVM con un algoritmo de entrenamiento de mínimos cuadrados y un kernel lineal consiguió una tasa de acierto del 99.19 % (media armónica igual a 99.12 % con una precisión del 99.42 % y una exhaustividad del 98.84 %). La solución para esta aplicación ya ha sido entregada a la empresa cliente, Microptic.

En segundo lugar, varias líneas de trabajo están abiertas para el desarrollo de un sistema de monitorización del desgaste de herramienta en fresadoras de bordes:

1. El método para la localización e identificación de plaquitas rotas que se basa en la información de varias capturas de imágenes de la misma plaquita bajo diferentes poses puede ser mejorado aumentando el número de plaquitas localizadas en cada imagen individual. Esto puede conseguirse utilizando, por ejemplo, una modificación de la transformada Hough para localizar elipses. Los tornillos tienen una forma circular cuando la plaquita se encuentra en el centro de la imagen, pero cuando la plaquita se encuentra en los lados de la imagen el tornillo es visto como una elipse. Otra opción es realizar la localización de plaquitas con filtros COSFIRE y después la localización de los filos de corte mediante detección de bordes y una transformada Hough. El disponer de más vista de una misma plaquita podría ayudar a mejorar los resultados ya que en algunas vistas hay un bajo contraste de la plaquita con respecto al fondo.

2. Los resultados también podrían mejorar con un sistema de iluminación más apropiado. Nuestro método confía en la detección de bordes y, por tanto, es necesario obtener un buen contraste. Una posible solución sería seguir a Pfeifer and Wiegers (2000) quienes capturaron la misma plaquita bajo diferentes posiciones de iluminación para combinar la información de todas las imágenes. De otro modo, se podrían evaluar diferentes configuraciones de iluminación y sistemas de captura.

3. Aunque la rotura de las plaquitas sea el aspecto más crítico a evaluar, sería interesante que el sistema de monitorización también evaluase el nivel de desgaste de las plaquitas. De este modo, podríamos predecir posibles rotura futuras de plaquitas o decidir cambiar plaquitas que alcancen altos niveles de desgaste.

Finalmente, el proyecto ASASEC ya ha sido entregado con resultados satisfactorios. La mayor parte del esfuerzo se puso en la programación apropiada del producto, siguiendo un diseño y análisis orientado a objetos y haciendo uso de patrones de diseño. Por consiguiente, decisiones futuras de cambios en la especificación del producto afectarán el programa de un modo más limitado. Sin embargo, las soluciones aportadas en el campo de la visión computacional consistieron principalmente en la implementación directa de métodos existentes, con un desarrollo limitado de nuevas técnicas de visión. Los filtros de color COSFIRE se pueden utilizar para muchas otras aplicaciones relacionadas con la recuperación y el reconocimiento de objetos en imágenes o vídeos de color. El principal inconveniente de los filtros de color COSFIRE es que requieren un alto número de convoluciones, que depende de la aplicación concreta, y que consumen tiempo. Sin embargo, la implementación de los filtros puede realizarse de modo paralelo o distribuido ya que la mayoría de los cálculos son independientes entre sí. Además, el enfoque de COSFIRE no se limita al uso de las respuestas de filtros Gabor. En el futuro, estudiaremos el reconocimiento de objetos utilizando una combinación de respuestas color SIFT (Van de Sande et al., 2010) en lugar de respuestas de filtros Gabor. Pruebas previas utilizando COSFIRE con SIFT para imágenes en escala de grises son prometedoras. Este enfoque puede resultar muy interesante para objetos que contienen *blobs* distintivos pero menos apropiado para objetos en los que los contornos sean determinantes.

Bibliography

Aldavert, D., Ramisa, A., de Mántaras, R. L. and Toledo, R.: 2010, Real-time object segmentation using a bag of features approach, *13th International Conference of the ACIA*, IOS Press, IOS Press, L'Espluga de Francolí, Catalonia, Spain.

Alegre, E., Biehl, M., Petkov, N. and Sánchez, L.: 2008, Automatic classification of the acrosome status of boar spermatozoa using digital image processing and LVQ, *Computers in Biology and Medicine* **38**(4), 461 – 468.

Alegre, E., Biehl, M., Petkov, N. and Sánchez, L.: 2013, Assessment of acrosome state in boar spermatozoa heads using n-contours descriptor and RLVQ, *Computer Methods and Programs in Biomedicine* **111**(3), 525 – 536.

Alegre, E., González-Castro, V., Alaiz-Rodríguez, R. and García-Ordás, M. T.: 2012, Texture and moments-based classification of the acrosome integrity of boar spermatozoa images, *Computer Methods and Programs in Biomedicine* **108**(2), 873 – 881.

Alegre, E., González-Castro, V., Suárez, S. and Castejón, M.: 2009, Comparison of supervised and unsupervised methods to classify boar acrosomes using texture descriptors, *ELMAR, 2009. ELMAR '09. International Symposium*, pp. 65–70.

Aller-Álvarez, N., Fernández-Robles, L., González-Castro, V. and Alegre, E.: 2015, Detección de plaquitas en un cabezal de fresado usando correspondencia de plaquitas, *Actas de las XXXVI Jornadas de Automática. Comité Español de Automática de la IFAC (CEA-IFAC)*, pp. 80–84.

Andreopoulos, A. and Tsotsos, J. K.: 2013, 50 years of object recognition: Directions forward, *Computer Vision and Image Understanding* **117**(8), 827 – 891.

Ans, B., Herault, J. and Jutten, C.: 1985, Adaptive neural architectures: detection of primitives, *Proc. of COGNITIVA*, pp. 593–597.

Atherton, T. and Kerbyson, D.: 1999, Size invariant circle detection, *Image and Vision Computing* **17**(11), 795 – 803.

Atli, A. V., Urhan, O., Ertürk, S. and Sönmez, M.: 2006, A computer vision-based fast approach to drilling tool condition monitoring, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture* **220**(9), 1409–1415.

Azzopardi, G. and Petkov, N.: 2013a, Automatic detection of vascular bifurcations in segmented retinal images using trainable cosfire filters, *Pattern Recognition Letters* **34**(8), 922–933.

Azzopardi, G. and Petkov, N.: 2013b, A shape descriptor based on trainable cosfire filters for the recognition of handwritten digits, *Computer Analysis of Images and Patterns - 15th International Conference, CAIP 2013, York, UK, August 27-29, 2013, Proceedings, Part II*, pp. 9–16.

Azzopardi, G. and Petkov, N.: 2013c, Trainable cosfire filters for keypoint detection and pattern recognition, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **35**(2), 490–503.

Azzopardi, G. and Petkov, N.: 2014, Ventral-stream-like shape representation: from pixel intensity values to trainable object-selective cosfire models, *Front. Comput. Neurosci.* **2014**.

Azzopardi, G., Strisciuglio, N., Vento, M. and Petkov, N.: 2015, Trainable cosfire filters for vessel delineation with application to retinal images, *Medical Image Analysis* **19**(1), 46–57.

Ballard, D.: 1981, Generalizing the Hough transform to detect arbitrary shapes, *Pattern Recognition* **13**(2), 111 – 122.

Barreiro, J., Castejón, M., Alegre, E. and Hernández, L.: 2008, Use of descriptors based on moments from digital images for tool wear monitoring, *International Journal of Machine Tools and Manufacture* **48**(9), 1005 – 1013.

Bartlett, M., Movellan, J. R. and Sejnowski, T.: 2002, Face recognition by independent component analysis, *Neural Networks, IEEE Transactions on* **13**(6), 1450–1464.

Baumberg, A.: 2000, Reliable feature matching across widely separated views, *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, Vol. 1, pp. 774–781 vol.1.

Bay, H., Ess, A., Tuytelaars, T. and Gool, L. V.: 2008, Speeded-up robust features (surf), *Computer Vision and Image Understanding* **110**(3), 346 – 359. Similarity Matching in Computer Vision and Multimedia.

Belongie, S., Malik, J. and Puzicha, J.: 2002, Shape matching and object recognition using shape contexts, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(4), 509–522.

Brummer, M.: 1991, Hough transform detection of the longitudinal fissure in tomographic head images, *Medical Imaging, IEEE Transactions on* **10**(1), 74–81.

Calonder, M., Lepetit, V., Strecha, C. and Fua, P.: 2010, Brief: Binary robust independent elementary features, *Proceedings of the 11th European Conference on Computer Vision: Part IV*, ECCV'10, Springer-Verlag, Berlin, Heidelberg, pp. 778–792.

Canny, J.: 1986, A computational approach to edge detection, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **8**(6), 679–698.

Carneiro, G. and Jepson, A.: 2003, Multi-scale phase-based local features, *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Vol. 1, pp. I–736–I–743 vol.1.

Castejón, M., Alegre, E., Barreiro, J. and Hernández, L.: 2007, On-line tool wear monitoring using geometric descriptors from digital images, *International Journal of Machine Tools and Manufacture* **47**(12-13), 1847 – 1853.

Chang, T. and Kuo, C.-C.: 1993, Texture analysis and classification with tree-structured wavelet transform, *Image Processing, IEEE Transactions on* **2**(4), 429–441.

Chethan, Y., Ravindra, H., Gowda, Y. K. and Kumar, G. M.: 2014, Parametric optimization in drilling EN-8 tool steel and drill wear monitoring using machine vision applied with taguchi method, *Procedia Materials Science* **5**(0), 1442 – 1449. International Conference on Advances in Manufacturing and Materials Engineering, ICAMME 2014.

Comon, P.: 1994, Independent component analysis, a new concept?, *Signal Processing* **36**(3), 287 – 314. Higher Order Statistics.

Dalal, N. and Triggs, B.: 2005a, Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 886–893.

Dalal, N. and Triggs, B.: 2005b, Histograms of oriented gradients for human detection, *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, Vol. 1, pp. 886–893 vol. 1.

Danesh, M. and Khalili, K.: 2015, Determination of tool wear in turning process using undecimated wavelet transform and textural features, *Procedia Technology* **19**(0), 98 – 105. 8th International Conference Interdisciplinarity in Engineering, INTER-ENG 2014, 9-10 October 2014, Tirgu Mures, Romania.

Datta, A., Dutta, S., Pal, S. and Sen, R.: 2013, Progressive cutting tool wear detection from machined surface images using voronoi tessellation method, *Journal of Materials Processing Technology* **213**(12), 2339 – 2349.

Dattner, I.: 2009, Statistical properties of the Hough transform estimator in the presence of measurement errors, *Journal of Multivariate Analysis* **100**(1), 112 – 125.

Dickinson, S. J., Leonardis, A., Schiele, B. and Tarr, M. J.: 2009, *Object Categorization, Computer and Human Vision Perspectives*, Cambridge University Pres.

Didion, B. A.: 2008, Computer-assisted semen analysis and its utility for profiling boar semen samples, *Theriogenology* **8**(70), 1374 – 1376.

Draper, B. A., Baek, K., Bartlett, M. S. and Beveridge, J.: 2003, Recognizing faces with PCA and ICA, *Computer Vision and Image Understanding* **91**(1-2), 115 – 137. Special Issue on Face Recognition.

Duda, R. O., Hart, P. E. and Stork, D. G.: 2000, *Pattern Classification (2nd Edition)*, Wiley-Interscience.

Dutta, S., Pal, S., Mukhopadhyay, S. and Sen, R.: 2013, Application of digital image processing in tool condition monitoring: A review, *CIRP Journal of Manufacturing Science and Technology* **6**(3), 212 – 232.

Ecabert, O., Peters, J., Schramm, H., Lorenz, C., von Berg, J., Walker, M., Vembar, M., Olszewski, M., Subramanyan, K., Lavi, G. and Weese, J.: 2008, Automatic model-based segmentation of the heart in ct images, *Medical Imaging, IEEE Transactions on* **27**(9), 1189–1201.

Favorskaya, M. and Proskurin, A.: 2015, Image categorization using color g-surf invariant to light intensity, *Procedia Computer Science* **60**, 681 – 690.

Fernández-Robles, L., Azzopardi, G., Alegre, E. and Petkov, N.: 2015, Cutting edge localisation in an edge profile milling head, *Computer Analysis of Images and Patterns*, Vol. 9257 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 336–347.

Francos, J., Meiri, A. and Porat, B.: 1993, A unified texture model based on a 2-D wold-like decomposition, *Signal Processing, IEEE Transactions on* **41**(8), 2665–2678.

Freeman, W. and Adelson, E.: 1991, The design and use of steerable filters, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **13**(9), 891–906.

García-Olalla, O., Alegre, E., Fernández-Robles, L., Malm, P. and Bengtsson, E.: 2015, Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors, *Computer Methods and Programs in Biomedicine* **120**(1), 49 – 64.

Gastal, E. S. L. and Oliveira, M. M.: 2011, Domain transform for edge-aware image and video processing, *ACM TOG* **30**(4), 69:1–69:12. Proceedings of SIGGRAPH 2011.

Golemati, S., Stoitsis, J., Balkizas, T. and Nikita, K.: 2005, Comparison of B-mode, M-mode and Hough transform methods for measurement of arterial diastolic and systolic diameters, *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, pp. 1758–1761.

González-Castro, V., Alegre, E., Morala-Argüello, P. and Suárez, S. A.: 2009, A combined and intelligent new segmentation method for boar semen based on thresholding and watershed transform, *International Journal of Imaging* **2**(S09), 70–80.

González, M., Alegre, E., Alaiz, R. and Sánchez, L.: 2007, Acrosome integrity classification of boar spermatozoon images using DWT and texture techniques, *VipIMAGE - Computational Vision and Medical Image Processing*, Algarve, Portugal, pp. 165–168.

Gordon, A. D.: 1999, *Classification*, Monographs on statistics and applied probability, Chapman & Hall, Boca Raton (Fla.), London, New York.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2003a, Contour detection based on nonclassical receptive field inhibition, *IEEE Transactions on Image Processing* **12**(7), 729–739.

Grigorescu, C., Petkov, N. and Westenberg, M. A.: 2003b, The role of non-CRF inhibition in contour detection, *The 11-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*.

Grigorescu, S., Petkov, N. and Kruizinga, P.: 2002, Comparison of texture features based on Gabor filters, *Image Processing, IEEE Transactions on* **11**(10), 1160–1167.

Guan, P. and Yan, H.: 2011, Blood cell image segmentation based on the Hough transform and fuzzy curve tracing, *Machine Learning and Cybernetics (ICMLC), 2011 International Conference on*, Vol. 4, pp. 1696–1701.

Guo, J., Shi, C., Azzopardi, G. and Petkov, N.: 2015, Recognition of architectural and electrical symbols by COSFIRE filters with inhibition, *Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II*, pp. 348–358.

Haar, A.: 1910, Zur theorie der orthogonalen funktionensystem, *Mathematische Annalen* **69**(3), 331–371.

Haralick, R. M.: 1979, Statistical and structural approaches to texture, *Proceedings of the IEEE* **67**(5), 786–804.

Harris, C. and Stephens, M.: 1988, A combined corner and edge detector, *In Proc. of Fourth Alvey Vision Conference*, pp. 147–151.

Hotelling, H.: 1936, Relations between two sets of variates, *Biometrika* **28**(3-4), 321–377.

Hough, P.: 1962, Method and Means for Recognizing Complex Patterns, U.S. Patent 3.069.654.

Huang, J., Kumar, S., Mitra, M., Zhu, W.-J. and Zabih, R.: 1997, Image indexing using color correlograms, *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pp. 762–768.

Hyvärinen, A. and Oja, E.: 2000, Independent component analysis: Algorithms and applications, *Neural Netw.* **13**(4-5), 411–430.

Illingworth, J. and Kittler, J.: 1988, A survey of the Hough transform, *Comput. Vision Graph. Image Process.* **44**(1), 87–116.

Jacobson, N., Nguyen, T. and Crosby, R.: 2007, Curvature scale space application to distorted object recognition and classification, *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, pp. 2110–2114.

Jagadish, H. V.: 1991, A retrieval technique for similar shapes, *Proceedings of the 1991 ACM SIGMOD International Conference on Management of Data*, SIGMOD '91, ACM, New York, NY, USA, pp. 208–217.

Jain, A. and Farrokhnia, F.: 1990, Unsupervised texture segmentation using Gabor filters, *Systems, Man and Cybernetics, 1990. Conference Proceedings., IEEE International Conference on*, pp. 14–19.

Johnson, A. E. and Hebert, M.: 1999, Using spin images for efficient object recognition in cluttered 3D scenes, *IEEE Trans. Pattern Anal. Mach. Intell.* **21**(5), 433–449.

Jolliffe, I.: 2002, *Principal Component Analysis*, Springer Series in Statistics, Springer.

Jurkovic, J., Korosec, M. and Kopac, J.: 2005, New approach in tool wear measuring technique using ccd vision system, *International Journal of Machine Tools and Manufacture* **45**(9), 1023 – 1030.

Jutten, C. and Herault, J.: 1991, Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture, *Signal Processing* **24**(1), 1 – 10.

Kadir, T., Boukerroui, D. and Brady, M.: 2003, An analysis of the scale saliency algorithm, *Technical report*, Robotics Research Laboratory, Department of Engineering Science.

Kadir, T. and Brady, M.: 2001, Saliency, scale and image description, *International Journal of Computer Vision* **45**(2), 83–105.

Kadir, T. and Brady, M.: 2003, Scale saliency: a novel approach to salient feature and scale selection, *Visual Information Engineering, 2003. VIE 2003. International Conference on*, pp. 25–28.

Kadir, T., Zisserman, A. and Brady, M.: 2004, An affine invariant salient region detector, *in* T. Pajdla and J. Matas (eds), *Computer Vision - ECCV 2004*, Vol. 3021 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 228–241.

Kalvoda, T. and Hwang, Y.-R.: 2010, A cutter tool monitoring in machining process using hilbert huang transform, *International Journal of Machine Tools and Manufacture* **50**(5), 495 – 501.

Kauppinen, H., Seppanen, T. and Pietikainen, M.: 1995, An experimental comparison of autoregressive and fourier-based descriptors in 2d shape classification, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **17**(2), 201–207.

Ke, Y. and Sukthankar, R.: 2004, Pca-sift: a more distinctive representation for local image descriptors, *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, Vol. 2, pp. II–506–II–513 Vol.2.

Kerr, D., Pengilley, J. and Garwood, R.: 2006, Assessment and visualisation of machine tool wear using computer vision, *The International Journal of Advanced Manufacturing Technology* **28**(7-8), 781–791.

Kim, J.-H., Moon, D.-K., Lee, D.-W., suk Kim, J., myung Chang Kang and Kim, K. H.: 2002, Tool wear measuring technique on the machine using ccd and exclusive jig, *Journal of Materials Processing Technology* **130-131**, 668–674.

Kiranyaz, S., Birinci, M. and Gabbouj, M.: 2010, Perceptual color descriptor based on spatial distribution: A top-down approach, *Image and Vision Computing* **28**(8), 1309 – 1326.

Kirby, M. and Sirovich, L.: 1990, Application of the karhunen-loeve procedure for the characterization of human faces, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12**(1), 103–108.

Koenderink, J. and van Doorn, A.: 1987, Representation of local geometry in the visual system, *Biological Cybernetics* **55**(6), 367–375.

Kruizinga, P. and Petkov, N.: 1999, Nonlinear operator for oriented texture, *Image Processing, IEEE Transactions on* **8**(10), 1395–1407.

Kurada, S. and Bradley, C.: 1997, A review of machine vision sensors for tool condition monitoring, *Computers in Industry* **34**(1), 55–72.

Laws, K. I.: 1979, Texture energy measures., *Proceedings Image Understanding Workshop*, pp. 47–51.

Lazebnik, S., Schmid, C. and Ponce, J.: 2003, A sparse texture representation using affine-invariant regions, *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, Vol. 2, pp. II–319–II–324 vol.2.

Lee, D. D. and Seung, H. S.: 1999, Learning the parts of objects by non-negative matrix factorization, *Nature* **401**(6755), 788–791.

Leutenegger, S., Chli, M. and Siegwart, R. Y.: 2011, BRISK: Binary robust invariant scalable keypoints, *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE, pp. 2548–2555.

Li, T., Mei, T., Kweon, I.-S. and Hua, X.-S.: 2011, Contextual bag-of-words for visual categorization, *Circuits and Systems for Video Technology, IEEE Transactions on* **21**(4), 381–392.

Li, Y.: 2005, *Object and Concept Recognition for Content-based Image Retrieval*, PhD thesis, University of Washington, Seattle, WA, USA. AAI3163394.

Liang, Y.-T., Chiou, Y.-C. and Louh, C.-J.: 2005, Automatic wear measurement of ti-based coatings milling via image registration., *MVA*, pp. 88–91.

Lim, T. and Ratnam, M.: 2012, Edge detection and measurement of nose radii of cutting tool inserts from scanned 2-d images, *Optics and Lasers in Engineering* **50**(11), 1628 – 1642.

Lindeberg, T.: 1998, Feature detection with automatic scale selection, *International Journal of Computer Vision* **30**(2), 79–116.

Liu, H., Feng, B. and Wei, J.: 2008, An effective data classification algorithm based on the decision table grid, *Computer and Information Science, 2008. ICIS 08. Seventh IEEE/ACIS International Conference on*, pp. 306–311.

Lowe, D.: 1999, Object recognition from local scale-invariant features, *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, Vol. 2, pp. 1150–1157.

Lowe, D.: 2004, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* **60**(2), 91–110.

Mair, E., Hager, G. D., Burschka, D., Suppa, M. and Hirzinger, G.: 2010, Adaptive and generic corner detection based on the accelerated segment test, *Proceedings of the 11th European Conference on Computer Vision: Part II*, ECCV'10, Springer-Verlag, Berlin, Heidelberg, pp. 183–196.

Makki, H., Heinemann, R., Hinduja, S. and Owodunni, O.: 2009, Online determination of tool run-out and wear using machine vision and image processing techniques, *Innovative Production Machines and Systems*.

Matas, J., Chum, O., Urban, M. and Pajdla, T.: 2004, Robust wide-baseline stereo from maximally stable extremal regions, *Image and Vision Computing* **22**(10), 761 – 767. British Machine Vision Computing 2002.

Matas, J. and Obdrzalek, S.: 2004, Object recognition methods based on transformation covariant features, *Signal Processing Conference, 2004 12th European*, pp. 1721–1728.

McManigle, J., Stebbing, R. and Noble, J.: 2012, Modified Hough transform for left ventricle myocardium segmentation in 3-d echocardiogram images, *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, pp. 290–293.

Mikolajczyk, K. and Schmid, C.: 2002, An affine invariant interest point detector, *Proceedings of the 7th European Conference on Computer Vision-Part I*, ECCV '02, Springer-Verlag, London, UK, UK, pp. 128–142.

Mikolajczyk, K. and Schmid, C.: 2004, Scale & affine invariant interest point detectors, *International Journal of Computer Vision* **60**(1), 63–86.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T. and Gool, L. V.: 2005, A comparison of affine region detectors, *Int. J. Comput. Vision* **65**(1-2), 43–72.

Nister, D. and Stewenius, H.: 2006, Scalable recognition with a vocabulary tree, *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, CVPR '06, IEEE Computer Society, Washington, DC, USA, pp. 2161–2168.

Ogiela, M. R. and Tadeusiewicz, R.: 2002, Syntactic reasoning and pattern recognition for analysis of coronary artery images, *Artificial Intelligence in Medicine* **26**(1-2), 145 – 159. Medical Data Mining and Knowledge Discovery.

Ogiela, M. and Tadeusiewicz, R.: 2005, Nonlinear processing and semantic content analysis in medical imaging - a cognitive approach, *Instrumentation and Measurement, IEEE Transactions on* **54**(6), 2149–2155.

Ojala, T., Pietikäinen, M. and Harwood, D.: 1996, A comparative study of texture measures with classification based on featured distributions, *Pattern Recognition* **29**(1), 51 – 59.

Otieno, A., Pedapati, C., Wan, X. and Zhang, H.: 2006, Imaging and wear analysis of microtools using machine vision, *IJME-INTERTECH International Conference*.

Ozuysal, M., Calonder, M., Lepetit, V. and Fua, P.: 2010, Fast keypoint recognition using random ferns, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(3), 448–461.

Paatero, P. and Tapper, U.: 1994, Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values, *Environmetrics* **5**(2), 111–126.

Pass, G. and Zabih, R.: 1996, Histogram refinement for content-based image retrieval, *Applications of Computer Vision, 1996. WACV '96., Proceedings 3rd IEEE Workshop on*, pp. 96–102.

Petkov, N.: 1995, Biologically motivated computationally intensive approaches to image pattern recognition, *Future Gener. Comput. Syst.* **11**(4-5), 451–465.

Petkov, N. and Kruizinga, P.: 1997, Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells., *Biological Cybernetics* **76**(2), 83–96.

Petkov, N. and Westenberg, M. A.: 2003, Suppression of contour perception by bandlimited noise and its relation to nonclassical receptive field inhibition, *Biological Cybernetics* **88**(3), 236–246.

Petkov, N. and Wieling, M.: 2008, Gabor filter for image processing and computer vision, *On line), http://matlabserver. cs. rug. nl/edgedetectionweb/index. html* .

Pfeifer, T. and Wiegers, L.: 2000, Reliable tool wear monitoring by optimized image and illumination control in machine vision, *Measurement* **28**(3), 209 – 218.

Prasad, K. N. and Ramamoorthy, B.: 2001, Tool wear evaluation by stereo vision and prediction by artificial neural network, *Journal of Materials Processing Technology* **112**(1), 43 – 52.

Rosten, E. and Drummond, T.: 2006, Machine learning for high-speed corner detection, *Proceedings of the 9th European Conference on Computer Vision - Volume Part I*, ECCV'06, Springer-Verlag, Berlin, Heidelberg, pp. 430–443.

Roth, P. M. and Winter, M.: 2008, Survey of Appearance-Based Methods for Object Recognition, *Technical report*, Institute for Computer Graphics and Vision, Graz University of Technology.

Rublee, E., Rabaud, V., Konolige, K. and Bradski, G.: 2011, Orb: An efficient alternative to sift or surf, *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2564–2571.

Sánchez, L., Petkov, N. and Alegre, E.: 2006, Statistical approach to boar semen evaluation using intracellular intensity distribution of head images, *Cellular and molecular biology* **52**(6), 38 – 43.

Schaffalitzky, F. and Zisserman, A.: 2002, Multi-view matching for unordered image sets, or "how do i organize my holiday snaps?", *in* A. Heyden, G. Sparr, M. Nielsen and P. Johansen (eds), *Computer Vision ? ECCV 2002*, Vol. 2350 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 414–431.

Schmid, C. and Mohr, R.: 1997, Local grayvalue invariants for image retrieval, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **19**(5), 530–535.

Schmid, C., Mohr, R. and Bauckhage, C.: 2000, Evaluation of interest point detectors, *Int. J. Comput. Vision* **37**(2), 151–172.

Shahabi, H. and Ratnam, M.: 2009, Assessment of flank wear and nose radius wear from workpiece roughness profile in turning operation using machine vision, *The International Journal of Advanced Manufacturing Technology* **43**(1-2), 11–21.

Shen, J. and Israël, G.: 1989, A receptor model using a specific non-negative transformation technique for ambient aerosol, *Atmospheric Environment (1967)* **23**(10), 2289 – 2298.

Shi, C., Guo, J., Azzopardi, G., Meijer, J. M., Jonkman, M. F. and Petkov, N.: 2015, Automatic differentiation of u- and n-serrated patterns in direct immunofluorescence images, *Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015 Proceedings, Part I*, pp. 513–521.

Shu, H., Luo, L., Bao, X., Yu, W. and Han, G.: 2000, An efficient method for computation of legendre moments, *Graphical Models* **62**(4), 237 – 262.

Sidibé, D., Sadek, I. and Mériaudeau, F.: 2015, Discrimination of retinal images containing bright lesions using sparse coded features and SVM, *Computers in Biology and Medicine* **62**, 175 – 184.

Silberberg, T. M., Davis, L. and Harwood, D.: 1984, An iterative Hough procedure for three-dimensional object recognition, *Pattern Recognition* **17**(6), 621 – 629.

Sivic, J. and Zisserman, A.: 2003, Video google: a text retrieval approach to object matching in videos, *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 1470–1477 vol.2.

Sivic, J. and Zisserman, A.: 2009, Efficient visual search of videos cast as text retrieval, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **31**(4), 591–606.

Smietanski, J., Tadeusiewicz, R. and Luczynska, E.: 2010, Texture analysis in perfusion images of prostate cancer - a case study, *Int. J. Appl. Math. Comput. Sci.* **20**(1), 149–156.

Smith, S. M. and Brady, J. M.: 1995, Susan - a new approach to low level image processing, *International Journal of Computer Vision* **23**, 45–78.

Sortino, M.: 2003, Application of statistical filtering for optical detection of tool wear, *International Journal of Machine Tools and Manufacture* **43**(5), 493 – 497.

Strisciuglio, N., Azzopardi, G., Vento, M. and Petkov, N.: 2015, Multiscale blood vessel delineation using B-COSFIRE filters, *Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015, Proceedings, Part II*, pp. 300–312.

Su, J., Huang, C. and Tarng, Y.: 2006, An automated flank wear measurement of microdrills using machine vision, *Journal of Materials Processing Technology* **180**(1-3), 328 – 335.

Sun, J., Sun, Q. and Surgenor, B.: 2012, An adaptable automated visual inspection scheme through online learning, *The International Journal of Advanced Manufacturing Technology* **59**(5-8), 655–667.

Tang, J., Wang, H. and Yan, Y.: 2015, Learning Hough regression models via bridge partial least squares for object detection, *Neurocomputing* **152**, 236 – 249.

Teague, M. R.: 1980, Image analysis via the general theory of moments, *Journal of the Optical Society of America* **70**(8), 920–930.

Tino, P., Zhao, H. and Yan, H.: 2011, Searching for coexpressed genes in three-color cdna microarray data using a probabilistic model-based Hough transform, *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* **8**(4), 1093–1107.

Tombari, F. and Di Stefano, L.: 2010, Object recognition in 3D scenes with occlusions and clutter by Hough voting, *Image and Video Technology (PSIVT), 2010 Fourth Pacific-Rim Symposium on*, pp. 349–355.

Tong, C. and Kamata, S.: 2010, 3D object matching based on spherical hilbert scanning, *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 2941–2944.

Turk, M. and Pentland, A.: 1991, Eigenfaces for recognition, *J. Cognitive Neuroscience* **3**(1), 71–86.

Tuytelaars, T. and Gool, L. V.: 1999, Content-based image retrieval based on local affinely invariant regions, *In Int. Conf. on Visual Information Systems*, pp. 493–500.

Tuytelaars, T. and Van Gool, L.: 2004, Matching widely separated views based on affine invariant regions, *Int. J. Comput. Vision* **59**(1), 61–85.

Van de Sande, K. E. A., Gevers, T. and Snoek, C. G. M.: 2010, Evaluating color descriptors for object and scene recognition, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **32**(9), 1582–1596.

Van Gool, L., Moons, T. and Ungureanu, D.: 1996, Affine / photometric invariants for planar intensity patterns, *in* B. Buxton and R. Cipolla (eds), *Computer Vision - ECCV '96*, Vol. 1064 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 642–651.

Vedaldi, A. and Fulkerson, B.: 2008, VLFeat: An open and portable library of computer vision algorithms, http://www.vlfeat.org/.

Veltkamp, R. and Hagedoorn, M.: 2001, State of the art in shape matching, *in* M. Lew (ed.), *Principles of Visual Information Retrieval*, Advances in Pattern Recognition, Springer London, pp. 87–119.

Wang, W., Hong, G. and Wong, Y.: 2006, Flank wear measurement by a threshold independent method with sub-pixel accuracy, *International Journal of Machine Tools and Manufacture* **46**(2), 199 – 207.

Wang, W., Wong, Y. and Hong, G.: 2005, Flank wear measurement by successive image analysis, *Computers in Industry* **56**(8-9), 816 – 830. Machine Vision Special Issue.

Xiong, G., Liu, J. and Avila, A.: 2011, Cutting tool wear measurement by using active contour model based image processing, *Mechatronics and Automation (ICMA), 2011 International Conference on*, pp. 670–675.

Yasmin, M., Mohsin, S. and Sharif, M.: 2014, Intelligent image retrieval techniques: A survey, *Journal of Applied Research and Technology* **12**(1), 87 – 103.

Yuen, H. K., Princen, J., Illingworth, J. and Kittler, J.: 1989, A comparative study of Hough transform methods for circle finding, *Proc. 5th Alvey Vision Conf., Reading (31 Aug*, pp. 169–174.

Zhang, C. and Zhang, J.: 2013, On-line tool wear measurement for ball-end milling cutter based on machine vision, *Computers in Industry* **64**(6), 708 – 719.

Zhang, J., Wu, J., Shi, X. and Zhang, Y.: 2012, Estimating vertebral orientation from biplanar radiography based on contour matching, *Computer-Based Medical Systems (CBMS), 2012 25th International Symposium on*, pp. 1–5.

Zhang, X., Tian, J., Deng, K., Wu, Y. and Li, X.: 2010, Automatic liver segmentation using a statistical shape model with optimal surface detection., *IEEE Trans. Biomed. Engineering* **57**(10), 2622–2626.

Zitnick, C. and Dollár, P.: 2014, Edge boxes: Locating object proposals from edges, *in* D. Fleet, T. Pajdla, B. Schiele and T. Tuytelaars (eds), *Computer Vision ECCV 2014*, Vol. 8693 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 391–405.

Zuiderveld, K.: 1994, Graphics gems iv, *in* P. S. Heckbert (ed.), *Contrast Limited Adaptive Histogram Equalization*, Academic Press Professional, Inc., San Diego, CA, USA, pp. 474–485.

Annex A: Research activities

Publications related with this manuscript

Evaluation of boar spermatozoa acrosomes

- L. Fernández-Robles, V. González-Castro, O. García-Olalla, M. T. García-Ordás and E. Alegre, *"A local invariant features approach for classifying acrosome integrity in boar spermatozoa,"* III Eccomas thematic conference on computational vision and medical image processing, VipIMAGE, Algarve, Portugal, October 12-14, ISBN-13:978-0-203-12818-3, 2011.
- L. Fernández-Robles, M. T. García-Ordás, D. García-Ordás, O. García-Olalla and E. Alegre, *"Acrosome evaluation of spermatozoa cells using sift and classical texture descriptors,"* XXXII Jornadas de Automática, Sevilla, Spain, September 7-9, ISBN:978-84-694-6454-0, 2011.
- L. Fernández-Robles, O. García-Olalla, M. T. García-Ordás, D. García-Ordás and E. Alegre, *"SVM approach to classify boar acrosome integrity of a multi-features SURF description,"* XXXIII Jornadas de Automática, Vigo, Spain, September 5-7, ISBN:978-84-8158-583-4, 2012, pp. 925-930.
- L. Fernández-Robles, E. Alegre, M. T. García-Ordás, O. García-Olalla, D. García-Ordás and E. Fidalgo, *"Combining SURF with global texture descriptors for classifying boar sperm,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.

Tool wear monitoring system for edge profile milling machine

- L. Fernández-Robles, G. Azzopardi, E. Alegre, and N. Petkov, *"Cutting edge localisation in an edge prole milling head,"* Computer Analysis of Images and Patterns - 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, Proceedings, Part II, Vol. 9257, pp. 336-347, 2015.
- L. Fernández-Robles, G. Azzopardi, E. Alegre, and N. Petkov, *"Machine-vision-based identification of broken inserts in edge profile milling heads,"* Submitted for publication.

- L. Fernández-Robles, G. Azzopardi, E. Alegre, N. Petkov and M. Castejón-Limas, *"Automatic identification of milling head inserts for early wear detection using trainable COSFIRE filters,"* Submitted for publication.

CBIR in ASASEC project

- L. Fernández-Robles, M. Castejón-Limas, J Alfonso-Cendón and E. Alegre, *"Evaluation of clustering configurations for object retrieval using SIFT features,"* Project Management and Engineering (selected papers from the 17th International AEIPRO Congress held in Logroño, Spain in 2013), Lecture Notes in Management and Industrial Engineering, Springer International Publishing, pp. 279-291, 2015.
- L. Fernández-Robles, J Alfonso-Cendón, M. Castejón-Limas, O. García-Olalla and E. Alegre, *"Development of an application for object retrieval and evaluation of the local invariant descriptors used,"* 19th International Congress on Project Management and Engineering, Granada, Spain, 15-17 July 2015.

Other publications

Evaluation of boar spermatozoa acrosomes

- O. García-Olalla, E. Alegre, L. Fernández-Robles, P. Malm and E. Bengtsson, *"Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors,"* Computer Methods and Programs in Biomedicine, Vol. 120, No. 1, pp. 49-64, 2015.
- O. García-Olalla, E. Alegre, L. Fernández-Robles and M. T. García-Ordás, *"Vitality assessment of boar sperm using an adaptive LBP based on oriented deviation,"* Computer Vision - ACCV 2012 Workshops, Lecture Notes in Computer Science, Vol. 7728, pp. 61-72, 2013.
- V. González-Castro, E. Alegre, O. García-Olalla, D. García-Ordás, M. T. García-Ordás and L. Fernández-Robles, *"Curvelet-based texture description to classify intact and damaged boar spermatozoa,"* Image Analysis and Recognition (ICIAR), Lecture Notes in Computer Science, vol. 7325, pp. 448-455, 2012.
- V. González-Castro, R. Alaiz-Rodríguez, L. Fernández-Robles, R. Guzmán-Martínez and E. Alegre, *"Estimating class proportions in boar semen analysis using the Hellinger distance,"* Trends in Applied Intelligent Systems (IEA/AIE), Lecture Notes in Computer Science, vol. 6096, pp. 284-293, 2010.
- M. T. García-Ordás, L. Fernández-Robles, O. García-Olalla, D. García-Ordás and E. Alegre, *"Boar spermatozoa classification using local invariant features and bag of words,"* XXXIII Jornadas de Automática, Vigo, Spain, September 5-7, ISBN:978-84-8158-583-4, pp. 947-952, 2012.

- E. Fidalgo, J. Pedro, L. Fernández-Robles, M. T. García-Ordás and E. Alegre, *"Evaluation of segmentation methods applied to intact and damaged boar spermatozoon heads,"* XXXIII Jornadas de Automática, Vigo, Spain, September 5-7, ISBN:978-84-8158-583-4, pp. 959-966, 2012.
- O. García-Olalla, M. T. García-Ordás, L. Fernández-Robles, D. García-Ordás and E. Alegre, *"Vitality assessment of boar sperm using N Concentric Squares resized and local binary pattern in gray scale images,"* XXXIII Jornadas de Automática, Vigo, Spain, September 5-7, ISBN:978-84-8158-583-4, pp. 919-924, 2012.

Co-author in the following inventions:

- Patent: Artificial vision procedure for the detection of proximal cytoplasmic droplets in spermatozoa[1], 2014.
- Patent: Artificial vision procedure for the detection of spermatozoa with curled tails[1], 2013.
- Intellectual property: Detection of proximal droplets in tails of spermatozoa using artificial vision techniques[1], 2011.
- Intellectual property: Detection of distal droplets in tails of spermatozoa using artificial vision techniques[1], 2011.
- Intellectual property: Classification of spermatozoa according to the state of their acrosomes using artificial vision techniques[1], 2011.
- Intellectual property: Detection of heads of spermatozoa with whip tails using artificial vision techniques[1], 2011.

Tool wear monitoring system for edge profile milling machine

- 0. García-Olalla, E. Alegre, J. Barreiro, L. Fernández-Robles and M T. García-Ordás, *"Tool wear classification using LBP-based descriptors combined with LOSIB-based enhancers,"* Proceedings of the 6th Manufacturing Engineering Society International Conference, Barcelona, Spain, 22-24 July, ISBN:978-84-1568863-1, pp. 174-182, 2015.
- 0. García-Olalla, L. Fernández-Robles, E. Fidalgo, V. González-Castro, and E. Alegre, *"Evaluation of the state of cutting tools according to its texture using LOSIB and LBP variants,"* 19th International congress on project management and engineering, Granada, Spain, 15-17 July 2015.
- A. M. de las Matas, V. González-Castro, L. Fernández-Robles and E. Alegre, *"Design and implementation of an embedded system for image acquisition of inserts in a headtool machine,"* XXXVI Jornadas de Automática, Bilbao, Spain, 2-4 September, ISBN:978-84-15914-12-9, pp. 147-152, 2015.

[1]Spanish Patent and Trademark Office, published in Spanish.

- N. Aller-Álvarez, L. Fernández-Robles, V. González-Castro and E. Alegre, *"Detección de plaquitas en un cabezal de fresado usando correspondencia de plantillas,"* XXXVI Jornadas de Automática, Bilbao, Spain, 2-4 September, ISBN:978-84-15914-12-9, pp. 80-84, 2015. Honoree.
- G. Martínez-San-Martín, L. Fernández-Robles, E. Alegre, and O. García-Olalla, *"A segmentation approach for evaluating wear of inserts in milling machines with computer vision techniques,"* XXXV Jornadas de Automática, Valencia, Spain, 3-5 September, ISBN:978-84-697-0589-6 , pp. 507-512, 2014.
- 0. García-Olalla, E. Alegre, J. Barreiro, L. Fernández-Robles and M T. García-Ordás, *"Tool wear classification using texture descriptors based on Local Binary Pattern,"* XXXV Jornadas de Automática, Valencia, Spain, 3-5 September, ISBN:978-84-697-0589-6 , pp. 292-298, 2014.

CBIR in ASASEC project

- N. Gorgojo, L. Fernández-Robles and E. Alegre, *"Object retrieval under different illumination directions using invariant local features,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.
- D. García-Ordás, L. Fernández-Robles, E. Alegre, M T. García-Ordás and O. García-Olalla, *"Automatic tampering detection in spliced images with different compression levels,"* Pattern Recognition and Image Analysis, Vol. 7887 of Lecture Notes in Computer Science, IbPRIA 2013, Madeira, Portugal, Springer Berlin Heidelberg, pp. 416-423, 2013.
- D. García-Ordás, L. Fernández-Robles, M T. García-Ordás, O. García-Olalla and E. Alegre, *"How to find an image despite it has been modified,"* Jornadas Nacionales de investigación en Ciberseguridad, León, Spain, September 2015.
- M-T. García-Ordás, O. García-Olalla, L. Fernández-Robles, D. García-Ordás and E. Alegre, *"Rotation invariant contour points descriptor histogram for shape based image retrieval,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.
- D. García-Ordás, E. Alegre, M T. García-Ordás, O. García-Olalla and L. Fernández-Robles, *"Robustness to rotation in perceptual hashing methods via dominant orientation,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.

Other topics

- O. García-Olalla, E. Alegre, L. Fernández-Robles and V. González-Castro, *"Local Oriented Statistics Information Booster (LOSIB) for Texture Classification,"* Pattern Recognition (ICPR 2014), 22nd International Conference on, 24-28 August, pp.1114-1119, 2014.
- O. García-Olalla, E. Alegre, L. Fernández-Robles, M T. García-Ordás and D. García-Ordás, *"Adaptive local binary pattern with oriented standard deviation (ALBPS) for texture*

Classification," EURASIP Journal on Image and Video Processing, Vol. 2013, No. 1, Springer International Publishing AG, 2013.

- O. García-Olalla, E. Alegre, M T. García-Ordás and L. Fernández-Robles, *"Evaluation of LBP Variants Using Several Metrics and kNN Classifiers,"* Similarity Search and Applications, Vol. 8199 on Lecture Notes in Computer Science, Springer Berlin Heidelberg, SISAP 2013, La Coruña, Spain, pp. 151-162, 2013.
- O. García-Olalla, M T. García-Ordás, L. Fernández-Robles, D. García-Ordás and E. Alegre, *"Comparison of different Local Binary Pattern variants for material recognition using KTH-TIPS 2a dataset,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.
- E. Fidalgo, L. Fernández-Robles, M T. García-Ordás, O. García-Olalla and E. Alegre, *"Evaluation of shape and color descriptors by using bag of Word techniques with one vs all classification,"* XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6, ISBN:978-84-616-5063-7, 2013.
- J. Alfonso-Cendón, M. Castejón-Limas and L. Fernández-Robles, *"Prototype development as tools for collaborative learning in projects Engineering in the context of the ESHE framework,"* Proceedings of the 17th International Congress on Project Management and Engineering, Logroño, Spain, pp. 1812-1820, 2013.
- J. Alfonso-Cendón, M. Castejón-Limas, G R. Delgado, I A. Juan and L. Fernández-Robles, *"iULE application for smart-phones for academic administration at the University of León,"* Proceedings of the 17th International Congress on Project Management and Engineering, Logroño, Spain, pp. 1712-1720, 2013.
- V. González-Castro, E. Alegre, O. García-Olalla, L. Fernández-Robles and M T. García-Ordás, *"Adaptive pattern spectrum image description using Euclidean and Geodesic distance without training for texture classification,"* Computer Vision, IET, Vol. 6, No. 6, pp. 581-589, November 2012.
- M.T. García-Ordás, L. Fernández-Robles, O. García-Olalla and D. García-Ordás, *"Words recognition using methods of word shape coding,"* XXXII Jornadas de Automática, Sevilla, Spain, September 7-9, ISBN:978-84-694-6454-0, 2011.
- D. García-Ordás, O. García-Olalla, L. Fernández-Robles and M T. García-Ordás, *"Video segmentation combining depth maps and intensity images,"* XXXII Jornadas de Automática, Sevilla, Spain, September 7-9, ISBN:978-84-694-6454-0, 2011.
- O. García-Olalla, D. García-Ordás, M T. García-Ordás and L. Fernández-Robles, *"Adaptive filters evaluation for sharpness enhancement and noise removal,"* XXXII Jornadas de Automática, Sevilla, Spain, September 7-9, ISBN:978-84-694-6454-0, 2011.
- E. Alegre, R. Alaiz-Rodríguez, J. Barreiro, E. Fidalgo and L. Fernández-Robles, *"Surface finish control in machining processes using Haralick descriptors and Neuronal Networks,"* Computational Modeling of Objects Represented in Images, Vol. 6026 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 231-241, 2010.

- S. A. Suárez-Castrillón, E. Alegre, J. Barreiro, P. Morala-Argüello and L Fernández-Robles, *"Surface roughness classification in metallic parts using Haralick descriptors and Quadratic Discriminant Analysis,"* Annals of DAAAM and Proceedings of the 21st International DAAAM Symposium, Vol. 21, No. 1, pp. 869-870, 2010.

Research projects

- Computer vision systems for life prediction of cutters for machining in severe conditions using fusion signals based classification. Spanish Ministry of Science and Innovation.
- ASASEC: Advisory System Against Sexual Exploitation of Children. European commission.
- Automatic assessment of fresh and criopreserved boar sperm through digital image segmentation, analysis and classification. Spanish Ministry of Science and Innovation.
- Proximal drops detection in boar spermatozoon tails through digital image processing. Microptic S.L.

Attended conferences

- 16th International Conference Computer Analysis of Images and Patterns, CAIP 2015, Valletta, Malta, September 2-4 2015.
- 19th International Congress on Project Management and Engineering, Granada, Spain, 15-17 July 2015.
- XXXIV Jornadas de Automática, Valencia, Spain, September 3-5 2014.
- XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6 2013.
- 11th Asian Conference on Computer Vision, ACCV 2012, Daejeon, Korea, November 5-9 2012.

Summer schools

- ICVSS, International Computer Vision Summer School, Siracusa, Sicily, 11-16th July 2011.
- INRIA, Visual Recognition and Machine Learning Summer School, Grenoble, France, 9-13 July 2012.

Annex B

SUMMARY OF THE THESIS IN SPANISH

RESUMEN DE LA TESIS EN CASTELLANO

En cumplimiento del punto 7° de la normativa complementaria del Real Decreto 778/1998, de 30 de Abril y de las normas para la aplicación del mismo, aprobadas por acuerdo de la Junta de Gobierno de fecha 10 de mayo de 1999, se adjunta un resumen en castellano de cada uno de los capítulos de esta tesis doctoral para que pueda admitirse a trámite.

1 Introducción

1.1 Motivación

El reconocimiento de objectos es una de las tareas fundamentales en visión computacional. Es el proceso de encontrar o identificar objetos en imágenes o vídeos digitales. Los métodos de reconocimiento de objectos generalmente usan la extracción de características y los algoritmos de aprendizaje para reconocer objetos o categorías de objetos. Aún hay importantes retos en el campo del reconocimiento de objetos. Una gran barrera la constituye la robustez en relación a la invarianza frente a escala, punto de vista, iluminación, deformaciones no rígidas y condiciones de captura de las imágenes. Otra es el gran escalado a miles de clases de objetos y millones de imágenes. En esta tesis consideramos de forma particular tres tareas del reconocimiento de objetos (Dickinson et al., 2009; Li, 2005):

- Clasificación: Dada una región de una imagen, decidir a qué categoría pertenece el objeto, u objetos presentes en dicha región.
- Detección y localización: Dada una imagen compleja, decidir si un objecto específico de interés está presente en algún lugar de esta imagen y proveer información precisa de su localización.
- Recuperación de imágenes mediante ejemplo (*content-based image retrieval* - CBIR): dada una imagen que típicamente contiene un objeto, imagen de ejemplo, y un conjunto de imágenes, recuperar las imágenes más similares a la de consulta devolviendo una lista de ellas ordenadas de mayor a menor parecido.

Esta tesis doctoral aborda las tareas antes descritas a través de tres aplicaciones diferentes: la clasificación de espermatozoides de verraco en función de la integridad de sus acrosomas, la localización de plaquitas montadas en un cabezal de fresado y el reconocimiento automático de las que presentan un filo de corte roto y finalmente la recuperación de objectos para el proyecto *Advisory System Against Sexual Exploitation of Children* (ASASEC) evaluando qué técnica de agrupamiento de puntos característicos es más apropiada y presentando un nuevo método para describir objetos de color.

1.1.1 Clasificación de espermatozoides de verraco en función de la integridad de sus acrosomas

Una mejor calidad de semen conduce a un mejor potencial de fertilización por inseminación artificial. La industria porcina desea obtener mejores ejemplares para su consumo. Los sistemas *Computer-Assisted Semen Analysis* (CASA) se utilizan para la evaluación de la calidad del semen (Didion, 2008). Sin embargo, no pueden analizar la integridad del acrosoma, que es un factor determinante. Habitualmente ésta se realiza de manera manual, con técnicas de tinción y conteo de los espermatozoides teñidos, requiriendo equipos de microscopía de alto coste. Hasta ahora los métodos automáticos que utilizan imágenes digitales adquiridas a través de microscopios en contraste de fase emplean técnicas estándar de descripción de textura de las cabezas de los espermatozoides. Estos enfoques necesitan segmentar apropiadamente dichas cabezas, que, en sí mismo, es un problema de difícil solución. Utilizando descriptores locales invariantes, *invariant local features* (ILF), esta segmentación se puede evitar. En esta tesis presentamos diversos métodos que permiten clasificar espermatozoides de verraco utilizando técnicas basadas en ILF.

1.1.2 Localización de las plaquitas rotas en cabezales de fresado en cabezales de bordes

La Fig. 1 muestra un cabezal de fresado que contiene herramientas de corte o plaquitas. La aplicación que aquí exponemos presenta dos desafíos: la localización de plaquitas y sus filos de corte; y la identificación de plaquitas rotas.

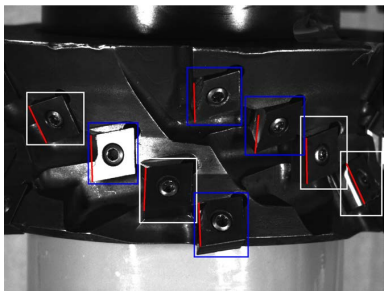


Figura 1: Cabezal de una fresadora de bordes. Los rectángulos blancos señalan plaquitas intactas y los azules rotas. Los segmentos rojos marcan los filos de corte ideales (intactos).

Los sistemas de monitorización de desgaste han presentado un gran desarrollado en las últimas décadas. Las plaquitas rotas constituyen una amenaza para la integridad del cabezal del corte (Kalvoda and Hwang, 2010). El tipo de fresado que nos incumbe se realiza en una sola pasada en condiciones agresivas. Es por ello que

las plaquitas pueden romperse sin sufrir un previo desgaste. El cambio de plaquitas es barato y rápido mientras que la rotura de un cabezal supone un elevado coste y un retraso en la producción. La localización de plaquitas es un reto porque el cabezal contiene 30 plaquitas siendo visibles entre 8 y 10 por imagen. Esta situación difiere de lo habitual con cabezales de dos plaquitas o disposiciones donde se puede capturar fácilmente una plaquita por imagen. Tecoi es una empresa interesada en este sistema y nos ha cedido el cabezal y las plaquitas para realizar este estudio.

1.1.3 Reconocimiento de objectos para la recuperación de imágenes mediante ejemplos

ASASEC es un proyecto europeo cuyo objetivo ha sidos proveer una solución tecnológica que ayude en la lucha contra la pornografía infantil. En este contexto, uno de los mayores retos consiste en la recuperación de imágenes y vídeos conteniendo objetos específicos y almacenados en grandes conjuntos de datos que provienen de casos previos. En el ámbito de este proyecto, se ha estudiado el reconocimiento de objetos para la recuperación de imágenes realizando consultas basadas en ejemplos de dichos objetos. Las imágenes recuperadas se ordenan en una lista en función de la similitud con el objeto utilizado como ejemplo en la consulta. El agrupamiento de características y la detección de objetos son dos tareas cruciales que se estudian en esta tesis doctoral. Se han comparado diferentes enfoques en el agrupamiento de características basados en una votación de la pose del objeto mediante transformada Hough y una verificación por mínimos cuadrados. Además, se ha añadido una descripción de color a los filtros *combination of shifted filter responses* (COSFIRE), que han obtenido muy buenos resultados previos en el reconocimiento de patrones lineales y de objetos en imágenes en escala de grises. De esta forma se ha conseguido una mejor descripción de los objetos que poseen formas similares y diferentes colores, así como un aumento de la eficiencia en el reconocimiento de objetos. Además, el método propuesto aporta invarianza a la intensidad del fondo del objeto.

1.2 Objetivos

El principal objetivo de esta tesis es la selección y evaluación de técnicas adecuadas de descripción y recuperación de objetos en diversas aplicaciones reales.

Dado dicho objetivo general se definen los siguientes objetivos particulares:

1. Evaluar la clasificación de espermatozoides de verraco de acuerdo a la integridad de sus acrosomas utilizando enfoques basados en ILF.

2. Proponer una solución automática para la identificación de plaquitas rotas en fresadoras de bordes que pueda ser instalada en línea sin retrasar las operaciones de mecanizado.

3. Estudiar dos campos específicos en el reconocimiento de objetos para CBIR en el ámbito de ASASEC: la evaluación de diferentes configuraciones de agrupaciones de características y la adición de descripción de color a los filtros COSFIRE.

1.3 Contribuciones principales

Las principales contribuciones de esta tesis doctoral pueden resumirse como se indica a continuación::

1. Se han utilizado ILF para la descripción del acrosoma de cabezas de espermatozoides de verraco. La clasificación de las cabezas como intactas y dañadas ha obtenido resultados satisfactorios. Los resultados de los métodos *speeded up robust features* (SURF) y *scale invariant feature transform* (SIFT) han mejorado a algunos descriptores globales de textura. Cuando este trabajo fue publicado como un artículo para una conferencia, estos eran los mejores resultados en la literatura.

2. Se ha realizado una propuesta que permite utilizar máquinas de vector de soporte (*support vector machine* - SVM) para trabajar con varios descriptores de características por imagen. Se ha presentado un método para clasificar descriptores SURF con SVM. La fusión de ambos tipos de descriptores ha proporcionado mejores resultados que los obtenidos individualmente por los mismos métodos.

3. Se ha presentado un método de fusión temprana de ILF con descriptores globales de textura para la clasificación de la integridad de los acrosomas de los espermatozoides de verraco. Este método ha mejorado a los métodos de forma individual.

4. Se ha propuesto un método altamente efectivo y eficiente para la localización de filos de plaquitas de corte montadas ens fresadoras. Su salida es un conjunto de regiones alrededor de los filos de corte, que puede ser utilizado como entrada para otros métodos que se centren en evaluar el estado del filo de corte.

5. Se ha introducido un método nuevo para la descripción efectiva y clasificación de las plaquitas con respecto al estado de sus filos de corte como rotas o

intactas. El tiempo que requiere para la inspección del cabezal de fresado es inferior al tiempo de reposo de la máquina.

6. Se ha presentado un nuevo método para la localización de plaquitas. Difiere del anterior porque considera cada imagen de forma individual y puede ser configurado de manera automática independientemente de la apariencia de las plaquitas. También se ha introducido una nueva métrica para el cálculo de la respuesta de filtros COSFIRE que mejora a las existentes previamente. Este método ha conseguido mejores resultados que trabajos previos basados en emparejamiento de plantillas o *template matching* (Aller-Álvarez et al., 2015).

7. Se han evaluado diversas configuraciones de agrupamientos de puntos clave SIFT en relación a sus parámetros de pose: localización de las coordenadas, escala y orientación. Por una parte se ha utilizado la medida de similitud del par de correspondencias más cercana. Por otro lado, se ha utilizado la transformada Hough, con diferentes parámetros, para identificar conjuntos de al menos tres puntos que voten por la misma pose de un objeto y se ha verificado su consistencia con el algoritmo de mínimos cuadrados.

8. Se ha propuesto un nuevo descriptor que tiene en cuenta tanto información de la forma como del color basado en los filtros COSFIRE (Azzopardi and Petkov, 2013c). A la vez de incorporar la descripción de color también añaden invarianza con respecto a la intensidad del fondo. Los filtros color COSFIRE se han presentado tanto para patrones hechos de líneas de color como para patrones que son objetos de color.

1.4 Organización del resto del documento

En el Capítulo 2 se expone un breve resumen de la revisión del estado de la técnica. En el Capítulo 3 se presentan métodos para la clasificación de espermatozoides de verraco en función de la integridad de su acrosoma basados en técnicas de ILF, junto con la experimentación y resultados obtenidos. El Capítulo 4 propone soluciones automáticas para la identificación de plaquitas rotas en fresadoras de bordes. Posteriormente, en el Capítulo 5, se describen las dos líneas de trabajo de reconocimiento de objetos para CBIR en el ámbito de ASASEC. Finalmente, el Capítulo 6 expone las conclusiones de esta tesis doctoral y las líneas futuras de trabajo.

2 Revisión del estado de la técnica

En las últimas décadas ha habido un trabajo sustancial en el campo de la visión computacional que aborda el problema del reconocimiento de objetos. En cuanto a la descripción del objeto, se pueden diferenciar métodos globales o locales según describan el objeto como un todo o localicen puntos o regiones de un objeto y describan parches del mismo.

La literatura engloba detectores locales basados en esquinas, regiones y otros. Los descriptores basados en esquinas localizan puntos clave o regiones que contienen estructuras tales como bordes. Las esquinas se pueden definir como puntos de baja similitud con sus vecinos en todas direcciones. El detector de esquinas más popular lo introdujo Harris and Stephens (1988). Este detector localiza un gran número de puntos clave con suficiente repetibilidad (Schmid et al., 2000). El detector Harris-Laplace añade invarianza a escala y se basa en el trabajo de Lindeberg (1998) que estudia las propiedades del espacio escala. Mikolajczyk and Schmid (2002) introdujo el detector Harris-Affine que extiende Harris-Laplace consiguiendo invarianza a transformaciones afines con el inconveniente de un aumento en el tiempo de cálculo. Los descriptores basados en regiones localizan regiones de luminosidad uniforme llamados *blobs* y son, por tanto, adecuados para regiones uniformes o para regiones con transiciones suaves. Los detectores Hessian (Mikolajczyk et al., 2005) son similares a los detectores Harris pero se basan en segundas derivadas, en lugar de primeras, por lo que el detector responde a estructuras de tipo *blob*. De igual modo, los descriptores Hessian-Laplace añaden invarianza a escala y los descriptores Hessian-Affine a transformaciones afines, (Mikolajczyk and Schmid, 2002). En lugar de una escala normalizada de Laplaciano, Lowe (1999, 2004) utiliza una aproximación de Laplaciano, denominada función de diferencia de Gaussianas (DoG), mediante el cálculo de imágenes emborronadas de Gaussianas con diferentes escalas locales contiguas. Regiones extremas de máximos estables (*maximally stable extremal regions* - MSER) (Matas et al., 2004) son regiones que son más oscuras o más claras que las que les rodean y que son estables a través de un rango de umbrales de la función intensidad. El número de regiones MSER detectadas es pequeño en comparación con los detectores previos pero Mikolajczyk et al. (2005) afirma que la repetibilidad es elevada en la mayoría de los casos. Otros descriptores están por ejemplo basados en regiones salientes de entropía (Kadir et al., 2003; Kadir and Brady, 2001, 2003), regiones de intensidad (Tuytelaars and Gool, 1999)y regiones de bordes (Tuytelaars and Van Gool, 2004). Algunos trabajos describen de forma local el objeto completo de manera densa, como por ejemplo la bolsa de palabras (*bag of words* - BoW) (Sivic and Zisserman, 2009) y los histogramas de gradientes orientados (*histogram of oriented gradients* - HOG) (Dalal and Triggs, 2005a).

Una vez obtenida una región de interés, esta debe ser descrita. Los descriptores basados en distribuciones representan algunas propiedades de la región mediante histogramas. Probablemente el descriptor más conocido es SIFT desarrollado por Lowe (1999, 2004). De hecho Lowe propuso de manera conjunta un detector de puntos clave, antes mencionado como DoG, y dicho descriptor. Se han propuesto muchas variantes de SIFT. Ke and Sukthankar (2004) redujo la dimensionalidad de SIFT mediante un análisis de componentes principales (PCA). Los histogramas de gradientes localización-componente (*gradient location-orientation histograms* - GLOH) tratan de obtener una mayor robustez y discriminación que SIFT, Mikolajczyk et al. (2005). Otros ejemplos son las imágenes en giro (Johnson and Hebert, 1999) o los descriptores cuyo contexto es la forma Belongie et al. (2002). En otra línea de investigación, los patrones locales binarios (LBP) introducidos por Ojala et al. (1996) son descriptores de textura basados en una codificación simple binaria de los valores de intensidad umbralizados. Este enfoque se ha extendido en muchas direcciones y campos con buenos resultados. Los descriptores basados en filtros capturan las propiedades de las regiones mediante el uso de filtros. Algunos métodos de prestigio son los descriptores diferenciales invariantes (Koenderink and van Doorn, 1987; Schmid and Mohr, 1997), los filtros dirigibles (Freeman and Adelson, 1991) o los filtros complejos (Baumberg, 2000; Carneiro and Jepson, 2003; Schaffalitzky and Zisserman, 2002). Otros descriptores se basan en el cálculo de momentos invariantes como los momentos de intensidad y color Van Gool et al. (1996) o los momentos gradiente Mikolajczyk et al. (2005).

Los métodos de descripción globales tradicionalmente se han basado en la descripción de la textura, la forma, el color, etc. del objeto. Otros métodos más complejos proyectan la imagen del objeto en un subespacio de inferior dimensión y se les conoce como métodos subespacio. Algunos de los más conocidos son PCA (Jolliffe, 2002; Turk and Pentland, 1991), factorización de matrices no negativa (NMF) (Paatero and Tapper, 1994; Shen and Israël, 1989; Lee and Seung, 1999), análisis de componentes independientes (ICA) (Ans et al., 1985; Hyvärinen and Oja, 2000; Comon, 1994; Jutten and Herault, 1991; Bartlett et al., 2002; Draper et al., 2003) o análisis de correlación canónica (CCA) (Hotelling, 1936).

Para una revisión más detallada, referimos al lector a (Andreopoulos and Tsotsos, 2013; Roth and Winter, 2008; Matas et al., 2004).

2.1 Clasificación de espermatozoides de verraco en función de la integridad de sus acrosomas utilizando ILF

Los trabajos existentes para la clasificación de la integridad de los acrosomas de espermatozoides de verraco están basados principalmente en descripción de tex-

tura. Siguiendo una línea temporal, González et al. (2007) utilizaron estadísticas de primer orden y características Haar en combinación con coeficientes Wavelet. La clasificación utilizando redes neuronales en un conjunto de 363 imágenes obtuvo una precisión del 92.19 %. Para el mismo conjunto de imágenes, Alegre et al. (2008) calcularon las magnitudes de los gradientes a lo largo de los contornos exteriores de las cabezas clasificando con cuatro prototipos un *learning vector quantization* (LVQ) llegando a precisiones del 93.2 %. Para 393 imágenes, Alegre et al. (2009) compararon características Haralick, máscaras Laws y momentos Zernike y Legendre clasificados con k vecinos más cercanos (k-nearest neighbours - k-NN) y análisis discriminante lineal (LDA) y cuadrático (QDA) consiguiendo tasas del 93.89 %.

Alegre et al. (2012) describieron los acrosomas mediante las estadísticas de primer orden derivadas de la matriz de co-ocurrencia de la imagen en niveles de gris, tanto en imagen original como en los coeficientes obtenidos por la transformada discreta Wavelet. Resultados experimentales en un conjunto de 800 imágenes obtuvieron un 94.93 % con un clasificador multicapa. Alegre et al. (2013) calcularon un descriptor de textura local para cada punto en siete contornos interiores de las cabezas. La clasificación con LVQ relevante consiguió una precisión del 99 % en un conjunto de 360 imágenes. El último trabajo, que conozcamos, es el desarrollado por García-Olalla et al. (2015) que combina descriptores de textura locales y globales y descriptores de contorno. La descripción global de textura se obtuvo de manera similar a Alegre et al. (2012) y la local por medio de LBP. Se utilizaron descriptores de forma Fourier para la caracterización del contorno. La concatenación de todos los descriptores y su clasificación con máquinas de vectores soporte (support vector machine - SVM) en 1851 imágenes obtuvo un 99.19 % de precisión.

No hemos encontrado ningún trabajo que se centre en la evaluación de la integridad de las células de esperma humanas o de cualquier animal utilizando enfoques basados en ILF. Los métodos basados en ILF se han hecho presentes en muchas tareas de reconocimiento de objetos desde que Lowe (1999) introdujo SIFT. El mayor problema de SIFT es el elevado tiempo de cálculo que precisa, por lo que muchas propuestas se han realizado tratando de mermar este inconveniente. Algunas son SURF (Bay et al., 2008), FAST (Rosten and Drummond, 2006), SUSAN (Smith and Brady, 1995), Ferns Ozuysal et al. (2010), BRIEF Calonder et al. (2010), BRISK (Leutenegger et al., 2011) and ORB (Rublee et al., 2011). A pesar de ello, SIFT y SURF siguen siendo los métodos más populares para el reconocimiento de objetos en las aplicaciones actuales.

2.2 Localización de plaquitas rotas en cabezales de fresado de bordes

Los sistemas de monitorización de desgaste (tool wear monitoring - TWM) han evolucionado mucho en las últimas décadas con el objetivo de evaluar el desgaste de herramientas de corte. Actualmente, el estado de la técnica presenta dos enfoques: directos e indirectos. El directo, el cual nos concierne, monitoriza el estado de la herramienta de corte directamente sobre el borde de corte cuando el cabezal se encuentra en la posición de descanso, Pfeifer and Wiegers (2000). El procesamiento de imágenes y la visión computacional han posibilitado un gran avance en las técnicas directas, Dutta et al. (2013).

Muchos trabajos que evalúan el estado de las plaquitas no realizan la tarea de su localización previa. Unos porque analizan directamente plaquitas desmontadas de manera individual (Castejón et al., 2007; Lim and Ratnam, 2012; Xiong et al., 2011). Otros porque utilizan cabezales con sólo dos plaquitas (Zhang and Zhang, 2013; Su et al., 2006; Kim et al., 2002) por lo que es sencillo realizar la captura de cada una de ellas. Y finalmente otros porque utilizan fresadoras de cara donde es fácil colocar el sistema de adquisición para capturar una sola plaquita por imagen (Jurkovic et al., 2005; Wang et al., 2006; Pfeifer and Wiegers, 2000; Sortino, 2003). Sin embargo, nuestra aplicación consta de un cabezal con 30 plaquitas por lo que cada imagen adquirida contiene entre 8 y 10 plaquitas, lo que hace que la localización de las mismas suponga un reto.

En cuanto a la identificación de plaquitas rotas, muchos son los enfoques que se han basado en el análisis de textura (Dutta et al., 2013). Algunos ejemplos se pueden leer en (Danesh and Khalili, 2015; Kerr et al., 2006; Barreiro et al., 2008; Datta et al., 2013; Prasad and Ramamoorthy, 2001). En cambio, el cabezal que aquí estudiamos realiza un fresado agresivo de gruesas placas en una sola pasada lo que puede causar la rotura de las plaquitas. Partes de la plaquita se rompen sin dañar la textura del resto de la misma, como se observa en los ejemplos de la Fig. 1. Por esta razón consideramos que la características basadas en textura, así como los ILF, no son adecuados para la aplicación que nos concierne.

Otros métodos utilizan la información de los contornos de las plaquitas para determinar su estado. Por ejemplo, Atli et al. (2006) clasificó herramientas de taladrado utilizando una medida de desviación de la linealidad de los bordes Canny detectados. Makki et al. (2009) detectó bordes y utilizó métodos de segmentación para describir el desgaste como una desviación de la porción del borde en taladros. Chethan et al. (2014) también comparó imágenes segmentadas de las plaquitas de taladros antes y después de su utilización en producción para determinar el estado de las mismas. De forma similar, Shahabi and Ratnam (2009) realizó un estudio com-

parando imágenes antes y después de su uso para torneado. En cuanto a fresado, los trabajos se centran en microfresado y fresado final. Así, por ejemplo, Otieno et al. (2006); Zhang and Zhang (2013); Liang et al. (2005) adquieren imágenes de las plaquitas antes y después de su uso para evaluar el estado de desgaste de las mismas. Todos los trabajos mencionados comparten un requisito común y es que necesitan adquirir una imagen de la plaquita intacta para evaluar las discrepancias con una nueva imagen de la misma herramienta.

Nosotros proponemos un nuevo algoritmo que evalúa el estado de las plaquitas sin necesidad de imágenes de referencia de plaquitas intactas. Esto evita la calibración del sistema cada vez que una plaquita es reemplazada y permite liberar memoria después de cada monitorización. Nuestro método determina automáticamente la posición y orientación de los filos de corte ideales en una imagen y calcula las desviaciones desde los filos de corte reales. Por tanto, con una sola imagen, se puede determinar si las plaquitas están rotas o no.

2.3 Reconocimiento de objectos para la recuperación de imágenes mediante ejemplos: transformada Hough y filtros COSFIRE para el reconocimiento de objetos

El reconocimiento de objetos para CBIR consiste en dos tareas difíciles: la identificación de los objetos en las imágenes y la búsqueda rápida de objetos en grandes colecciones de imágenes. En esta tesis nos centramos en la primera tarea. Hay muchos trabajos dedicados al reconocimiento de objetos para CBIR, unos basados en texturas (Chang and Kuo, 1993; Francos et al., 1993; Jain and Farrokhnia, 1990; Smietanski et al., 2010), formas (Jagadish, 1991; Kauppinen et al., 1995; Veltkamp and Hagedoorn, 2001), colores representation (Huang et al., 1997; Kiranyaz et al., 2010; Pass and Zabih, 1996) o descripción de contornos (Ogiela and Tadeusiewicz, 2002, 2005; Zitnick and Dollár, 2014). Recientemente, las técnicas de ILF han ganado popularidad (Lowe, 2004; Matas and Obdrzalek, 2004,?; Mikolajczyk and Schmid, 2004; Nister and Stewenius, 2006; Sivic and Zisserman, 2003).

La transformada Hough (Hough, 1962) puede utilizarse para el reconocimiento de objetos. Aunque originalmente se definió para identificar formas simples, como líneas y círculos, su uso ha sido extendido a formas más generales, permitiendo la detección de múltiples instancias de un objeto que puede estar parcialmente ocluido. Ballard (1981) introdujo la transformada de Hough generalizada que modifica la primera haciendo uso del principio del emparejamiento de plantillas. De esta forma la transformada Hough puede ser utilizada para detectar cualquier objeto descrito por su modelo. Al utilizar una transformada Hough, se puede crear un esquema de votación como es común en la aplicación de este algoritmo (Illingworth and Kittler,

1988). La transformada Hough cubre inmensas aplicaciones prácticas en las tareas de reconocimiento de objetos. Lowe (2004) la utilizó para identificar conjuntos de descriptores de características que pertenecen a un único objeto mediante una votación de cada característica a todos las poses del objeto que son consistentes con la misma. Tang et al. (2015) introdujo un nuevo esquema de votación multi-escala en el que múltiples imágenes Hough correspondientes a múltiples escalas del objeto pueden ser obtenidas simultáneamente para hacer frente a los cambios de escala del objeto. En la detección de objetos en tres dimensiones, Silberberg et al. (1984), Tombari and Di Stefano (2010) y Tong and Kamata (2010) utilizaron la transformada Hough satisfactoriamente. La transformada Hough también se ha utilizado con éxito en imágenes médicas para el reconocimiento de objetos, como en (Golemati et al., 2005; McManigle et al., 2012; Ecabert et al., 2008; Zhang et al., 2010; Guan and Yan, 2011; Tino et al., 2011).

Los filtros COSFIRE fueron propuestos por Azzopardi and Petkov (2013c) para la localización de patrones locales que consistían en una combinación de segmentos de contorno. Se han aplicado exitosamente en aplicaciones para la delineación de vasos sanguíneos (Azzopardi et al., 2015; Strisciuglio et al., 2015), bifurcaciones vasculares (Azzopardi and Petkov, 2013a), reconocimiento de dígitos escritos a mano (Azzopardi and Petkov, 2013b) o la diferenciación de patrones lineales típicos en enfermedades de piel (Shi et al., 2015). Su efectividad para el reconocimiento de objetos también ha sido demostrada. Azzopardi and Petkov (2013c) aplicó filtros COSFIRE para el reconocimiento de tres tipos de señales de señales consiguiendo resultados perfectos de detección y reconocimiento en un conjunto de 48 imágenes de tráfico. Un modelo jerárquico y entrenable de reconocimiento de objetos se propuso para la aplicación de un robot que recoge zapatos (Azzopardi and Petkov, 2014) permitiendo detectar objetos deformables en escenas complejas sin necesidad de una segmentación previa. Guo et al. (2015) introdujo inhibición a los filtros COSFIRE para que únicamente respondan a la combinación de contornos presentada en la configuración y no a combinaciones de lo anterior con otros segmentos de contorno. Aplicaron estos filtros para el reconocimiento de símbolos arquitectónicos y eléctricos demostrando la efectividad del método incluso en imágenes con ruido. Sin embargo, que conozcamos, no hay trabajos que utilicen la información de color para mejorar la detección de objetos de color ni tampoco que provea una solución para la invarianza a la intensidad del fondo de imagen.

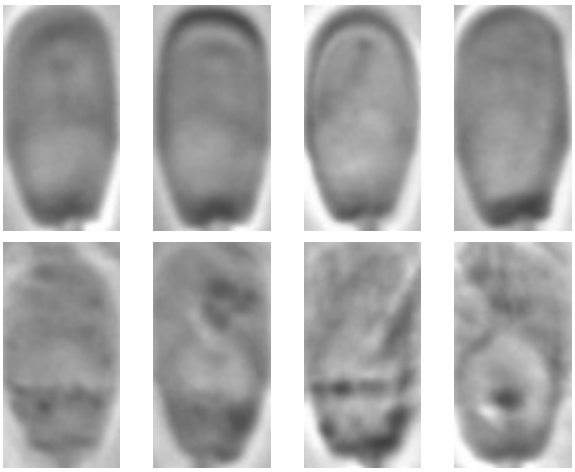


Figura 2: (arriba) Acrosomas registrados intactos. (abajo) Acrosomas registrados dañados.

3 Clasificación de espermatozoides de verraco en función de la integridad de sus acrosomas

3.1 Conjunto de imágenes

Se ha utilizado una cámara Basler Scout scA780-54fc y un microscopio de fluorescencia Nikon E-600. Este microscopio permite visualizar imágenes tanto en contraste de fases como en fluorescencia. Se han tomado dos imágenes por muestra de semen, una en contraste de fases para evaluar la metodología propuesta y otra en fluorescencia para realizar el etiquetado y comparar resultados. Las imágenes han sido adquiridas con una resolución de 780×580 píxeles y un aumento de 100× del microscopio. Por tanto, no más de tres o cuatro cabezas aparecen en cada captura. Como la mayoría de espermatozoides provienen de diferentes capturas, la iluminación no es completamente constante. Para cada imagen las cabezas de espermatozoides han sido recortadas. Las cabezas que presentaban solapamientos han sido descartadas. Cada cabeza ha sido automáticamente registrada para asegurar invarianza a escala y rotación. Finalmente las imágenes mal registradas han sido manualmente descartadas. El conjunto se compone de 856 imágenes de cabezas intactas y 861 imágenes de cabezas dañadas de 56×108 píxeles. La Fig. 2 muestra ejemplos de cabezas de espermatozoides registradas tanto intactas como dañadas.

3.2 Características locales invariantes frente a descriptores de textura tradicionales

En esta sección mostramos la comparación de los resultados de métodos basados en ILF y descriptores globales tradicionales de textura para la clasificación de cabezas de espermatozoides de verraco como intactas o dañadas.

3.2.1 Método

Hemos extraído 13 de las 14 características propuestas por Haralick (1979) de la imagen original, todas menos el coeficiente de correlación máximo siguiendo a (Alegre et al., 2009). Hemos calculado la matriz de co-ocurrencia en escala de grises (*gray-level co-occurrence matrix* - GLCM) para cuatro direcciones $\theta = \{0^o, 45^o, 90^o, 135^o\}$ calculando la media para lograr invarianza a rotación. Además, hemos calculado la GLCM para las distancias $d = \{1, 2, 3, 5\}$.

Por otro lado, hemos aplicado la transformada discreta de ondícula o wavelet (DWT) a las imágenes mediante wavelet de tipo Haar (Haar, 1910). Hemos calculado de nuevo las mismas 13 características de Haralick en la GLCM de la imagen original y en las GLCM de las cuatro sub-imágenes del primer nivel de descomposición de las DWT de Haar, obteniendo un descriptor compuesto por 65 características que denominamos WCF13.

Además hemos utilizado las máscaras de Laws (Laws, 1979). Hemos utilizado los vectores de nivel (L_5), bordes (E_5), manchas (S_5) y ondas (R_5) tal y como recomienda Alegre et al. (2009), obteniendo 16 máscaras calculando la convolución los pares de vectores. Primero hemos normalizado las imágenes sustrayendo de cada píxel la media de sus 15×15 vecinos para eliminar efectos de iluminación. Después, hemos realizado la convolución de cada imagen con las máscaras de Laws y se obtienen los mapas de energía mediante una media móvil no lineal de los valores absolutos. Finalmente, hemos combinado los mapas de energía de pares de filtros simétricos dados, consiguiendo 9 descriptores que son: L_5E_5/E_5L_5, L_5S_5/S_5L_5, L_5R_5/R_5L_5, E_5E_5, E_5S_5/S_5E_5, E_5R_5/R_5E_5, S_5S_5, S_5R_5/R_5S_5 and R_5R_5.

También se han obtenido los momentos Legendre Shu et al. (2000) y Zernike Teague (1980). Hemos utilizado 9 momentos de Legendre, desde el orden (0,0) al orden (2,2) y los valores absolutos de 9 momentos de Zernike, hasta el cuarto orden.

Como métodos de ILF hemos utilizado SIFT (Lowe, 2004) y SURF (Bay et al., 2008) con descriptores de 128 y 64 características respectivamente.

3.2.2 Experimentación

Hemos clasificado las imágenes de test mediante k-NN, tomando valores impares de k comprendidos entre 1 y 15 incluidos. SIFT y SURF obtienen muchos descriptores por imagen. Hemos calculado las distancias de cada descriptor en la imagen de test a todos los descriptores en la imagen de entrenamiento y consideramos la correspondencia que consigue la mínima distancia. Repetimos este cálculo para todos los descriptores de la imagen de test y calculamos la suma de esas distancias mínimas que se considera la distancia entre las imágenes de entrenamiento y test. Después utilizamos k-NN. La proximidad entre los patrones se ha calculado mediante la distancia Euclídea para todos los métodos utilizados y además hemos utilizado la similitud coseno para SIFT tal y como proponen sus autores (Lowe, 2004). Tomamos un conjunto aleatorio del 70 % de las imágenes de cada clase para el entrenamiento y el resto para el test. Consideramos un verdadero positivo (VP) cuando un acrosoma dañado se clasifica como tal, un falso positivo (FP) cuando un acrosoma intacto se clasifica como dañado, un falso negativo (FN) cuando un acrosoma dañado se clasifica como intacto y un verdadero negativo (VN) cuando un acrosoma intacto se clasifica como intacto. Calculamos la tasa de acierto como la tasa de clasificaciones correctas sobre todo el conjunto de test tasa de acierto $= (\text{VP} + \text{VN})/(\text{VP} + \text{FP} + \text{FN} + \text{VN})$. Como el número de imágenes en las clases son similares esta medida es correcta. Este proceso se repite 10 veces y la tasa de acierto final se calcula como la media de las tasas en los 10 conjuntos. También se ha calculado la tasa de acierto para cada clase, intactos y dañados.

3.2.3 Resultados

La Tabla 1 muestra las tasas de acierto para cada método evaluado para el número de vecinos k que obtuvo mejores resultados. La tasa de acierto más alta se consiguió con SURF y $k = 11$ (94.88 %), mejorando al resto de descriptores en todos los valores de k utilizados. Los resultados SIFT son similares independientemente de la métrica utilizada, distancia Euclídea o similitud coseno; aunque la distancia Euclídea siempre consiguió resultados ligeramente mejores. SURF y SIFT obtuvieron mejores tasas de acierto para la clase de dañados que para la clase de intactos para todos los valores de k evaluados. SURF consiguió una tasa de acierto igual a 92.89 % en la clase de intactos mientras que llegó al 96.86 % para los dañados. Por el contrario, los descriptores de textura obtuvieron mejores tasas de acierto para la clase de intactos con excepción de Haralick. Consideramos que una combinación de descriptores de textura global y de ILF puede mejorar los resultados individuales.

Tabla 1: Tasas de acierto de cada método evaluado para el número de vecinos que consiguió una mayor tasa de acierto global.

Descriptores	**k**	**Global (%)**	**Intacto (%)**	**Dañado (%)**
SURF	11	**94,88**	**92,89**	**96,86**
Legendre	7	87,55	88,24	86,86
Laws	7	86,95	91,98	81,94
SIFT Euclidea	11	84,64	76,15	92,96
SIFT cosine	11	84,24	75,52	92,80
WCF13 d=3	15	74,76	74,90	74,61
Zernike	11	74,46	86,80	62,21
Haralick d=2	1	71,22	70,39	72,05

3.3 Clasificación con SVM de descriptores SURF

3.3.1 Motivación

Muchos algoritmos de clasificación comúnmente conocidos sólo admiten un vector descriptor por imagen. Por este motivo, los métodos basados en ILF no pueden utilizar directamente dichos clasificadores. Generalmente, los trabajos que usan ILF confían en algoritmos de vecinos más cercanos para clasificar los descriptores de los puntos clave (Lowe, 2004). Nosotros hemos adaptado el algoritmo SVM para lidiar con varios vectores descriptores por imagen. Algunos trabajos construyen histogramas (BoW) a partir de los descriptores de ILF para obtener un vector de tamaño fijo por imagen que luego pueden usar en SVM (Sidibé et al., 2015; Favorskaya and Proskurin, 2015). Sin embargo, nuestro objetivo es utilizar los descriptores sin la necesidad de agruparlos.

3.3.2 Método y experimentación

Primero, concatenamos todos los descriptores SURF de todas las imágenes en una matriz de 17122×64, donde cada fila representa un descriptor y cada columna una característica SURF. Además, definimos un vector etiqueta de 17122 elementos en el cual cada punto perteneciente a una cabeza intacta se etiqueta como intacto y viceversa.

Realizamos dos experimentos. En primer lugar, clasificamos únicamente puntos clave de las cabezas de espermatozoides. Entrenamos un SVM basado en un algoritmo de mínimos cuadrados. Llevamos a cabo una validación cruzada de k interaciones, o k-folds, con $k = 10$ para todos los puntos clave. Calculamos las tasas

de acierto como el ratio de puntos clave correctamente clasificados sobre todo el conjunto de puntos clave. Finalmente, calculamos la media de los resultados para las 10 divisiones. En segundo lugar, clasificamos las cabezas de espermatozoides. De nuevo utilizamos un SVM para descriptores individuales, pero esta vez realizamos la validación cruzada de 10-fold en las cabezas y no en los puntos clave. Esto significa que los descriptores pertenecientes al 90 % de las cabezas son seleccionados en cada interación sin importar cuántos puntos clave contengan dichas cabezas. Además, ahora se calcula el resultado en términos de clasificación de cabezas y no de puntos clave. Consideramos una cabeza como correctamente clasificada cuando obtiene un número más alto de puntos bien clasificados por SVM que erróneamente clasificados. Calculamos la tasa de acierto como el ratio de cabezas correctamente clasificadas en el total de cabezas. Finalmente, se obtiene la media de los resultados en las 10 iteraciones.

3.3.3 Resultados

La Fig. 3 muestra las tasas de acierto con los dos enfoques propuestos. Las tasas de acierto para cada clase también se muestran. Se obtuvo una tasa de acierto de tan solo 72.57 % clasificando puntos clave mientras que llegó a 90.91 % al clasificar cabezas, lo que supone una mejora del 25.27 %. Nos gustaría resaltar, que al clasificar puntos clave las cabezas dañadas obtuvieron mejores tasas de acierto que las intactas, mientras que al clasificar cabezas sucede lo contrario. Puede deberse a que las cabezas dañadas tienen puntos clave que son más distintivos que los intactos pero las cabezas dañadas contienen áreas donde se visualiza que el acrosoma está dañado junto con otras en las que no se aprecia.

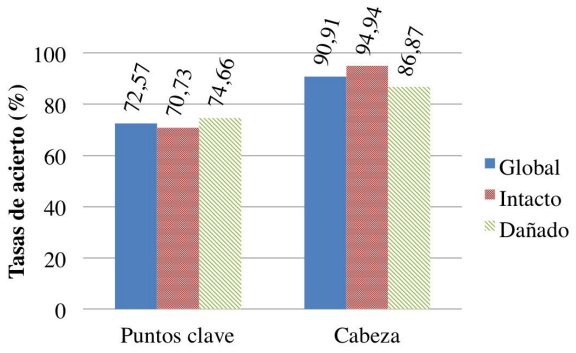


Figura 3: Tasas de acierto globales, para la clase intacta y para la clase dañada, utilizando SURF y SVM aplicado a puntos clave y cabezas.

Este enfoque puede extenderse a diferentes ILF (SIFT, BRISK, FREAK, etcétera) y a otros algoritmos de clasificación convencionales como redes neuronales.

3.4 Combinación de ILF y descriptores de textura globales

3.4.1 Método

Hemos combinado descriptores SIFT y SURF con momentos de Legendre y Zernike, máscaras de Laws y características Haralick ya explicados en la Sección 3.2.1. En cada imagen, realizamos una fusión temprana de los descriptores ILF con cada descriptor de textura global. Hemos elegido concatenar todos los descriptores de ILF de una imagen con el descriptor de textura global obtenido para la misma imagen. Por tanto, el emparejamiento se ve afectado directamente por la dimensionalidad de los descriptores originales. Antes de la fusión, se han normalizado los descriptores individuales a media 0 y desviación estándar 1.

3.4.2 Experimentos

Hemos utilizado los algoritmos k-NN y SVM para realizar la clasificación. k-NN se implementa como se indica en la Sección 3.2.2. SVM se aplica mediante un modelo de BoW (Aldavert et al., 2010; Li et al., 2011). Empleamos BoW con un algoritmo de agrupamiento k-means con $k = \{2, 3, ..., 10\}$ y distancia Euclídea para asignar la proximidad entre descriptores. Los histogramas de BoW se utilizan para entrenar un conjunto de imágenes de entrenamiento con SVM y un algoritmo lineal de mínimos cuadrados. Finalmente, calculamos las tasas de acierto obtenidas en la clasificación del conjunto de test. Definimos tasa de acierto del mismo modo que en la Sección 3.2.2. Tomamos aleatoriamente un 70 % de las imágenes de cada clase para entrenamiento y el resto para test. El proceso se repite 10 veces y los resultados son promediados.

3.4.3 Resultados

La Fig. 4 muestra los resultados obtenidos con la fusión temprana de ILF y descriptores de textura globales. Los mejores resultados se consiguieron al combinar ILF con momentos de Legendre y clasificando con k-NN. La mejor tasa de acierto global de 95.56 % se obtuvo con la combinación de SURF y Legendre y de 88.98 % con SIFT y Legendre. Estas combinaciones mejoran los resultados conseguidos individualmente por cada método con el mismo algoritmo de clasificación (94.88 %, 84.86 % y 87.55 % para SURF, SIFT and Legendre respectivamente) como se presentó en la Sección 3.2.3. SVM con BoW obtuvo resultados inferiores. El bajo número de

puntos clave detectados en estas imágenes de baja resolución (56×108 píxeles) puede ser la causa de una pobre definición del diccionario. Por tanto no se recomienda el uso de BoW para esta aplicación.

4 Localización automática de plaquitas rotas en fresadoras de bordes

4.1 Conjunto de imágenes

Hemos creado un conjunto de imágenes del cabezal de una fresadora que hemos hecho público[2]. Se compone de 144 imágenes de un cabezal de bordes de una fresadora. El cabezal tiene forma cilíndrica y contiene 30 plaquitas en total, de las cuales de 7 a 10 plaquitas resultan visibles en cada imagen. En el cabezal hay 6 grupos de 5 plaquitas colocadas diagonalmente a lo largo de la dirección axial de su perímetro. La última de las plaquitas de cada grupo está alineada verticalmente con la primera del siguiente grupo. En total hay por tanto 24 posiciones diferentes a lo largo del perímetro radial del cabezal en las que una plaquita está alineada con la cámara a intervalos de 15°. Por ello, la misma plaquita se captura en diferentes imágenes (entre 7 y 9) bajo diferentes poses cuando el cabezal es rotado, Fig. 5. La evaluación de plaquitas se realiza durante el estado de reposo de la fresadora entre el procesamiento de dos placas metálicas. El sistema de captura puede ser montado en esta posición de reposo.

Creamos el conjunto de imágenes siguiendo un proceso iterativo. Montamos 30 plaquitas en el cabezal y tomamos 24 imágenes rotando el mismo a intervalos de 15°. Repetimos este proceso 6 veces con diferentes plaquitas, recopilando en total 144 imágenes que contienen 180 plaquitas únicas de las cuales 19 están rotas. Utilizamos una cámara Genie M1280 $1/3''$ con resolución 1280×960 píxeles y una lente AZURE-2514MM. El cabezal fue iluminado con dos barras LED BDBL-R(IR)82/16H. Junto con las imágenes, se presentan las máscaras de los filos de corte ideales etiquetados con un número identificador único. También se adjunta un etiquetado de cada filo de corte como roto o no roto. En la Fig. 5 presentamos tres imágenes consecutivas que contienen las mismas plaquitas en diferentes posiciones debido a la rotación del cabezal. Además, creamos unas máscaras alrededor de cada tornillo de 40 píxeles de radio que cubren prácticamente los tornillos como se muestra en la segunda fila de la Fig. 5.

[2] http://pitia.unileon.es/varp/node/395

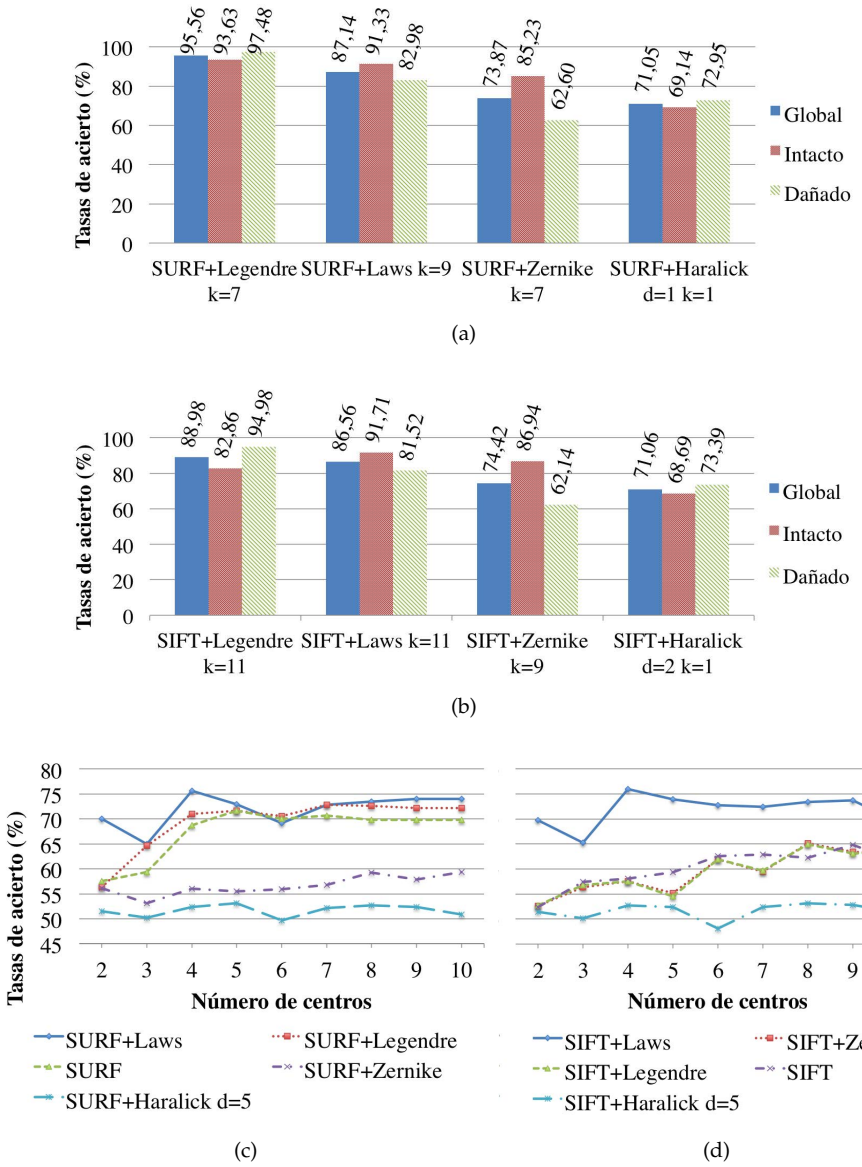


Figura 4: Resultados de la fusión temprana de ILF y descriptores globales de textura. (a) SURF como ILF y clasificación mediante k-NN. (b) SIFT como ILF y clasificación mediante k-NN. (c) SURF como ILF y clasificación mediante BoW y SVM. (d) SIFT como ILF y clasificación mediante BoW y SVM.

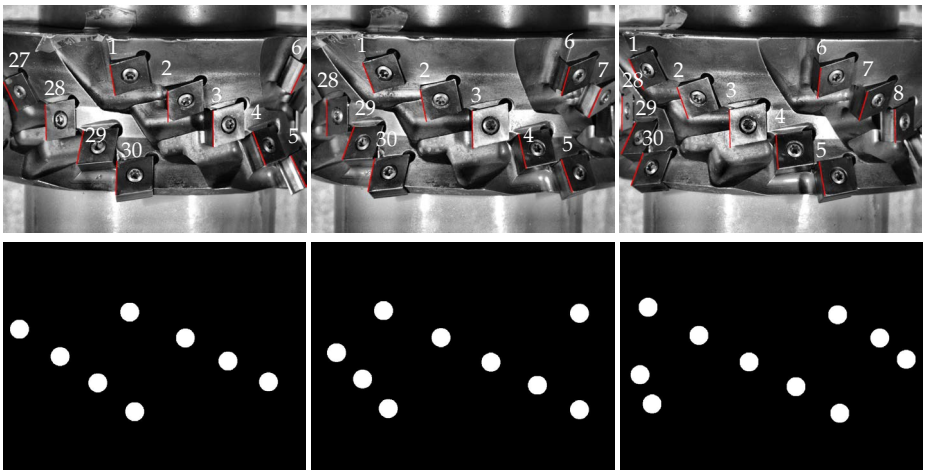


Figura 5: En la primera fila, los números indican el etiquetado de cada filo de corte en tres imágenes consecutivas del conjunto de imágenes. En la segunda fila, las máscaras circulares están situadas en los centros de los tornillos. Los círculos blancos cubren aproximadamente los tornillos de la imagen.

4.2 Localización automática de plaquitas y filos de corte utilizando procesamiento de imágenes

4.2.1 Método

Primero detectamos los tornillos que sujetan las plaquitas y a continuación localizamos los filos de corte. Para mejorar la calidad de las imágenes, en primer lugar hemos aplicado un método de ecualización de histograma de contraste limitado (CLAHE) Zuiderveld (1994).

Los tornillos que sujetan las plaquitas tienen una forma circular. Utilizamos la transformada circular Hough (CHT) para detectar círculos con radios entre 20 y 40 píxeles, porque es el tamaño de un tornillo en las imágenes de 1280×960 píxeles. Se ha utilizado un algoritmo en dos etapas para el cálculo del acumulador de la CHT Atherton and Kerbyson (1999) Yuen et al. (1989). Recortamos un área rectangular de tamaño 205×205 píxeles centrada en el tornillo detectado. Esas dimensiones son suficientes para contener una plaquita entera sin importar en qué posición del cabezal se encuentre. Después utilizamos este área recortada para identificar el filo de corte.

Las plaquitas tienen una forma romboidal formada por dos filos casi verticales ($\pm$ 22^o) y dos filos casi horizontales ($\pm$ 20^o). En primer lugar, utilizamos el método

Canny (Canny, 1986) para detectar bordes en el área recortada (Fig. 6(a-b)). Después aplicamos una transformada de Hough estándar (SHT) (Hough, 1962) a la imagen de bordes para detectar líneas rectas, Fig. 6c y elegimos las líneas rectas más probables de formar los filos de la plaquita en base a la geometría de la herramienta y las condiciones fijas del entorno de la cámara respecto al cabezal. Las líneas se consideran infinitas y se muestran en la Fig. 6d. Los puntos donde las líneas horizontales intersectan a la línea vertical izquierda definen los límites del filo de corte. Estos puntos se han marcado en azul en la Fig. 6d y el filo de corte localizado en la Fig. 6e. Si una de las líneas no se detecta, utilizamos la simetría para determinar la misma. Los tres últimos ejemplos en la Fig. 6 muestran esta situación. El método presentado generaliza la detección de filos de corte incluso en situaciones de filos rotos o desgastados. Finalmente, definimos una región de interés (ROI) mediante la dilatación del filo de corte detectado con un elemento estructural cuadrado de 10 píxeles de lado.

4.2.2 Experimentos y resultados

Si el filo de corte verdadero (etiquetado manualmente por expertos) se sitúa completamente en la ROI, contamos la ROI como un éxito y cuando no se sitúa en ninguna ROI, el resultado de éxito es 0. Si el filo de corte verdadero se solapa con la ROI, el resultado de éxito es igual a la fracción del filo de corte verdadero que se sitúa dentro de la ROI. Algunos ejemplos se muestran en la Fig. 7.

Cada plaquita se detecta en al menos una de las 144 imágenes. Además, cuando una plaquita es detectada, el filo de corte siempre es obtenido. Calculamos la precisión del método como la media de los resultados parciales para los filos de corte individuales. Usando este protocolo, se obtuvo una precisión del 99.61 %.

4.2.3 Discusión

Según nuestro conocimiento, este es el primer método que localiza automáticamente múltiples plaquitas y filos de corte en una fresadora de bordes. Los parámetros utilizados se han calculado de manera que son generales para todas las plaquitas independientemente de su posición en el cabezal dada su geometría y el posicionamiento del sistema de captura. Para una fresadora específica, los parámetros pueden ser estimados fácilmente y después no es necesario ningún otro ajuste. En el futuro, las ROI estimadas pueden ser utilizadas para evaluar el estado del filo de corte. Además, el método propuesto puede ser utilizado para diferentes cabezales que contengan plaquitas poligonales sujetas con tornillos, como es el diseño típico de las fresadoras. Hemos implementado el método en Matlab y hemos ejecutado los experimentos en un ordenador personal de 2 GHz de procesador y 8 GB de RAM.

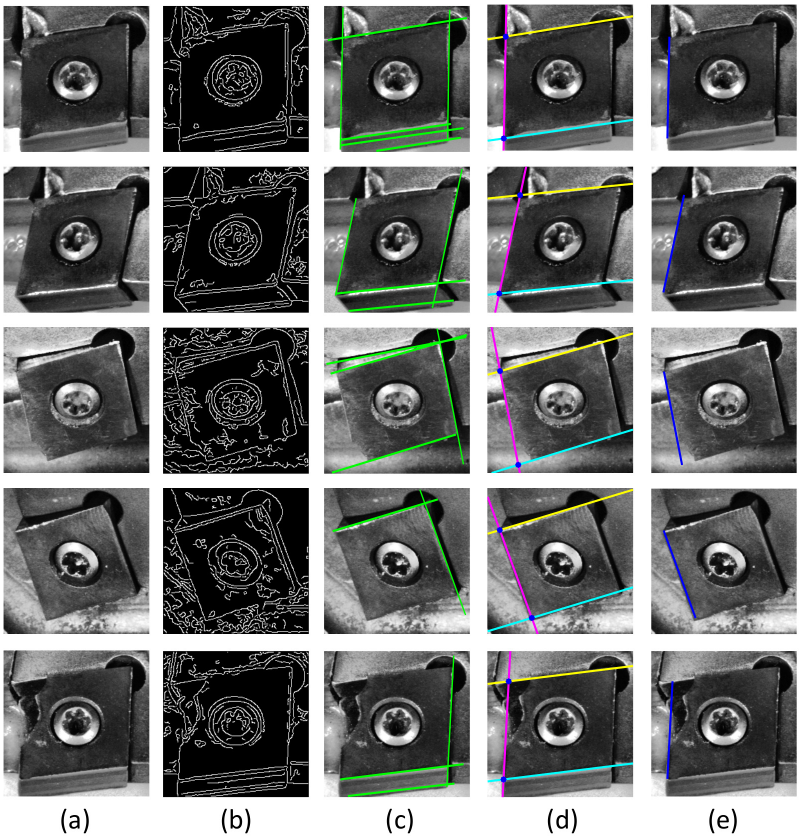
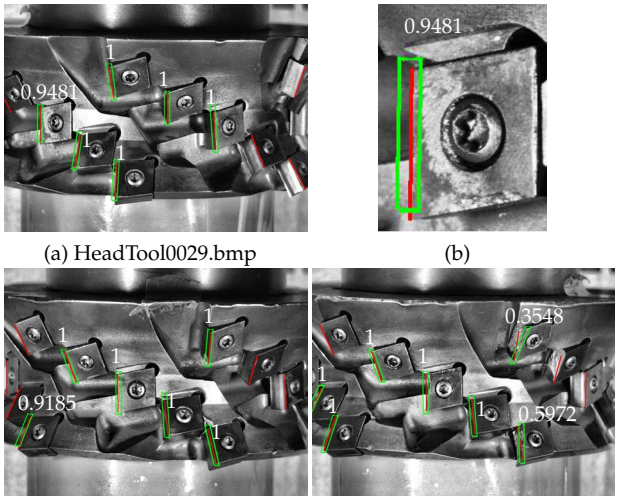


Figura 6: (a) Área recortada que contiene una plaquita. (b) Mapas de bordes Canny. (c) Detección de líneas (casi) verticales y (casi) horizontales. (d) Los puntos azules marcan la intersección entre las dos líneas horizontales y la línea vertical izquierda. (e) Filos de corte detectados.

El procesamiento de todos los pasos en una imagen requiere menos de 1,5 segundos y se requiere alrededor de 1 minuto para la captura y procesamiento de las 24 imágenes tomadas al cabezal. Las fresadoras con las que se hizo el estudio están en la posición de descanso entre 5 y 30 minutos, por lo que la implementación realizada podría ejecutarse en tiempo real.



(a) HeadTool0029.bmp (b)

(c) HeadTool0047.bmp (d) HeadTool0139.bmp

Figura 7: Los cuadriláteros verdes son las ROI detectadas por el método propuesto y las líneas rojas representan los filos de corte verdaderos. Los valores de éxito de las plaquitas se han indicado en fuente blanca.

4.3 Clasificación de plaquitas como rotas o no rotas

4.3.1 Método

Consideramos una plaquita no rota la que tiene un filo de corte recto, de lo contrario se considera rota. En primer lugar localizamos las plaquitas y los filos de corte intactos como se explicó en la sección anterior. Después, evaluamos las plaquitas mediante un método en tres pasos: aplicamos un filtro de suavizado que preserva bordes, calculamos el gradiente de cada filo y finalmente usamos las propiedades geométricas de los filos para evaluar su estado.

Determinamos una ROI a partir de los filos de corte ideales localizados con el método expuesto en la sección anterior, Fig. 8a. Una ROI está determinada por dos líneas paralelas al filo de corte ideal, una de 3 píxeles a la izquierda del mismo y otra a una distancia de 0,7 veces el espacio entre el filo de corte ideal y el centro del tornillo, w. Además consideramos una línea paralela al filo superior 3 píxeles hacia abajo y otra línea paralela al filo inferior 3 píxeles hacia arriba. Del cuadrilátero resultante eliminamos el segmento circular (de radio 45 píxeles) alrededor del tornillo que coincide con el cuadrilátero. Esta ROI permite evaluar el estado del filo de corte a la vez que ignora posibles bordes desgastados de los filos inferior y superior del tornillo. Finalmente, consideramos un área rectangular alrededor de la ROI con un

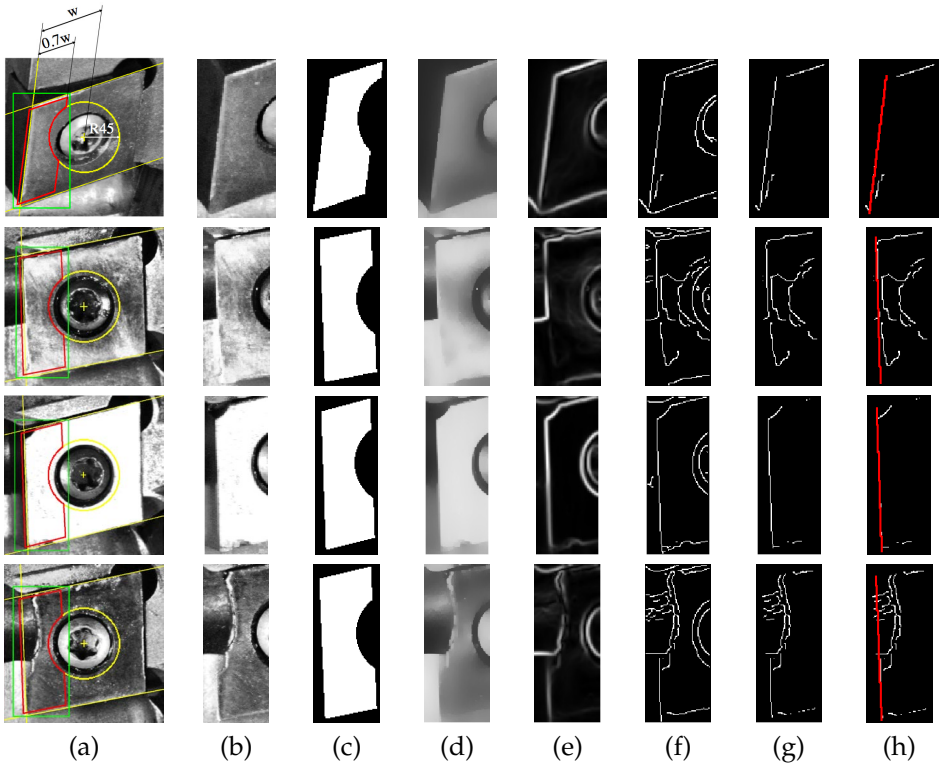


Figura 8: (a) En amarillo, los filos de corte ideal, superior e inferior. En rojo, la definición de la ROI. En verde, el área rectangular a recortar. (b) Región recortada. (c) Máscara que define la ROI en el área rectangular. (d) Región filtrada con conservación de bordes. (e) Mapa de magnitudes gradiente. (f) Mapa de bordes. (g) Resultado de multiplicar el mapa de bordes por la máscara. (h) Filo de corte real en blanco y filo de corte ideal en rojo. Las dos plaquitas superiores no están rotas mientras que las dos inferiores están rotas.

borde de 3 píxeles y la utilizamos para recortar la región de la imagen que contienen la ROI, Fig. 8b. También consideramos una máscara que define la ROI en dicha área rectangular, Fig. 8c.

La textura heterogénea y el bajo contraste de la plaquita con respecto al cabezal hacen de la detección de los filos de corte reales una tarea compleja. En primer lugar, aplicamos el filtro de suavizado que conserva los bordes propuesto por Gastal and Oliveira (2011) a la región recortada, Fig. 8d. A continuación, aplicamos el método Canny (Canny, 1986) para obtener el mapa de magnitudes gradiente, Fig. 8e, y bordes, Fig. 8f. Finalmente, consideramos únicamente los píxeles comprendidos en la ROI como los puntos que definen el filo de corte real, Fig. 8g. Para cada píxel del filo

de corte real conocemos sus coordenadas y el valor de su magnitud gradiente.

Para cada punto del filo de corte ideal determinamos el conjunto de puntos del filo de corte real que se sitúan en una línea paralela al filo superior detectado y que pasa por el punto del filo de corte ideal considerado. A continuación, calculamos las distancias Euclídeas del punto del filo de corte ideal considerado a cada punto del filo de corte real que cumple las condiciones anteriores. Y obtenemos la información de la mínima distancia y de la magnitud del gradiente del punto del filo de corte real que cumple esta mínima distancia. Considerando todos los puntos del filo de corte ideal, obtenemos un vector de desviaciones del filo de corte real dado por el conjunto de mínimas distancias así como un vector de magnitudes gradiente. También calculamos la media de este vector de magnitudes gradiente y la consideramos como el valor de magnitud gradiente de la ROI.

Eliminamos desviaciones anormales que habitualmente son causadas por determinadas texturas de la superficie de la plaquita. Para ello, aplicamos un filtro de mediana y eliminamos desviaciones que superan un cierto umbral respecto a los valores medianos. Este proceso de eliminación de espurios se realiza dos veces. Además eliminamos desviaciones cuando sus correspondientes magnitudes gradiente son inferiores a un umbral global dado. Esto se debe a que consideramos que el borde no está lo suficientemente definido como para evaluar el estado del filo de corte en esa zona. Para asegurar que una plaquita está rota, la desviación debe ser suficientemente alta a lo largo de una región del bode de corte y no únicamente en un píxel aislado. Para ello, aplicamos un filtro de media al vector de desviaciones resultante. Por último, tomamos el valor máximo después de aplicar el filtro de media como el valor de desviación de la ROI considerada.

En definitiva, cada ROI está representada por su valor de desviación y por su valor de magnitud gradiente.

Recordamos al lector que la misma plaquita es detectada en varias imágenes bajo diferentes poses. Para cada plaquita, calculamos la desviación y la magnitud gradiente en cada ROI detectada de dicha plaquita. Clasificamos una plaquita como rota si la imagen con la mayor magnitud gradiente presenta una desviación mayor de un umbral dado T, o si las desviaciones de al menos dos ROI (independientemente de sus magnitudes gradiente) son mayores que T. De otro modo, clasificamos la plaquita como no rota.

4.3.2 Experimentos y resultados

Utilizamos Matlab en un ordenador personal con procesador de 2 GHz y RAM de 8 GB. El proceso completo de identificación de plaquitas rotas en un cabezal con 30 plaquitas tarda menos de 3 minutos. Este tiempo es suficiente para la apli-

cación ya que de acuerdo con los expertos consultados la fresadora permanece en la posición de reposo entre 5 y 30 minutos, durante los cuales la chapa fresada es reemplazada por una nueva.

Nuestro conjunto de imágenes es sesgada con 19 plaquitas rotas y 161 no rotas. Nos referimos a las plaquitas rotas como la clase positiva y las no rotas como la negativa. Por tanto, un verdadero positivo (VP) es una plaquita rota clasificada como tal; un falso positivo (FP) es una plaquita no rota clasificada como rota y un falso negativo (FN) es una plaquita rota clasificada como no rota. Calculamos la precisión $P = VP/(VP + FP)$, exhaustividad $R = VP/(VP + FN)$ y la media armónica $F = 2PR/(P + R)$ para un conjunto de umbrales $T \in \{5; 5{,}01; \ldots; 8\}$ y obtenemos la curva $P - R$. Consideramos el mejor par (P, R), el que contribuye a obtener la máxima media armónica.

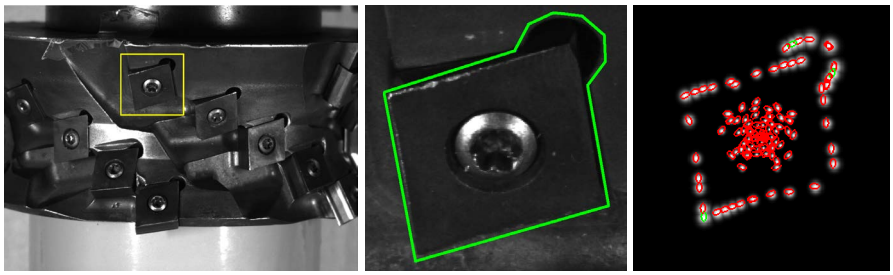
Aplicamos una validación repetida de sub-muestras aleatorias donde en cada iteración dividimos aleatoriamente el conjunto de imágenes en los subconjuntos de entrenamiento ($70\,\%$) y validación ($30\,\%$). Para cada iteración, utilizamos los datos de entrenamiento para determinar el conjunto de parámetros que consigue una media armónica máxima global F. Se obtiene aplicando una búsqueda por combinaciones y calculando el máximo valor de la media armónica de cada combinación. Si varias combinaciones de parámetros obtienen la misma media armónica, tomamos una cualquiera de forma aleatoria. El conjunto de parámetros determinado es utilizado posteriormente para evaluar el conjunto de validación. Repetimos el proceso 20 veces y finalmente promediamos los resultados obtenidos en los conjuntos de validación. Hemos obtenido un promedio de media armónica de $F = 0{,}9143(\pm 0{,}079)$ con una precisión igual a $P = 0{,}9661(\pm 0{,}073)$ y una exhaustividad del $R = 0{,}8821(\pm 0{,}134)$.

4.4 Localización automática de plaquitas utilizando COSFIRE

4.4.1 Método

Este método que aquí presentamos, con respecto al anterior utilizando procesamiento de imágenes, considera cada imagen del data set de manera independiente y puede ser configurado sin utilizar ninguna información previa sobre la apariencia de las plaquitas. La Fig. 9 muestra un ejemplo del conjunto de imágenes. En amarillo se encuentra enmarcada una plaquita, que seleccionamos como primer prototipo para calcular un filtro COSFIRE.

El filtro COSFIRE se calcula siguiendo (Azzopardi and Petkov, 2013c). En este caso, los bordes más característicos se encuentran en los bordes de la plaquita, alrededor del tornillo y en la grieta superior del cabezal. Utilizamos filtros de Gabor antisimétricos $\zeta = \pi/2$ con los siguientes parámetros: el ratio de aspecto $\gamma = 0{,}3$, el



(a) (b) (c)

Figura 9: (a) Imagen de entrada. En amarillo la región de interés del prototipo. (b) Región de interés ampliada que se utiliza para configurar el filtro COSFIRE y selección de la ROI. (c) Estructura del filtro COSFIRE configurado. Cada elipse representa una tupla del conjunto de las partes de contornos. Las longitudes de onda y orientaciones de los filtros Gabor utilizados se han tenido encuenta en la representación. Las manchas brillantes son mapas de intensidad de las funciones Gaussianas que se usan en la aplicación para emborronar las respuestas de los filtros Gabor.

ancho de banda de 1,5 por lo que $\sigma = 0{,}39\lambda$ y una longitud de onda $\lambda = 6$. Además hemos considerado todas las respuestas de los filtros Gabor, $t_1 = 0$.

Para configurar el filtro COSFIRE, tomamos 25 radios ρ a distancias iguales entre 0 y la mitad de la diagonal de la región de interés que engloba al prototipo y un umbral para mantener únicamente respuestas fuertes a lo largo de las circunferencias consideradas de $t_2 = 0,15$.

La aplicación de un filtro COSFIRE configurado para una plaquita en una imagen de entrada también se realiza siguiendo el método original descrito en (Azzopardi and Petkov, 2013c). La respuesta $r_{S_f}(x,y)$ de un filtro COSFIRE se define como una función de las respuestas de filtros Gabor emborronadas y desplazadas para las partes de contorno descritas en la configuración. Evaluamos tres funciones diferentes: media aritmética (AM) -Eq. 1-, media geométrica estricta (HGM) -Eq. 2- y media geométrica suave (SGM) -Eq. 3-.

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \frac{1}{|S_f|} \left(\sum_{i=1}^{|S_f|} (s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)) \right) \right|_{t_3} \tag{1}$$

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \left(\prod_{i=1}^{|S_f|} (s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)) \right)^{1/|S_f|} \right|_{t_3} \tag{2}$$

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left| \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) + \epsilon \right) \right)^{1/|S_f|} \right|_{t_3} \tag{3}$$

donde $|.|_{t_3}$ es la respuesta después de realizar una umbralización a fracciones t_3 del máximo global. Consideramos que los puntos que obtienen respuesta vienen dados por los máximos locales de la repuesta al filtro COSFIRE que se encuentren separados al menos una distancia de 200 píxeles. Por la forma del cabezal y el posicionamiento del sistema de captura cualquier par de plaquitas se encuentran separadas por una distancia mayor de 200 píxeles. Consideramos estos puntos como verdaderos positivos (VP) si se encuentran dentro de las coordenadas del etiquetado, de otro modo los consideramos como falsos positivos (FP).

La respuesta de un filtro COSFIRE utilizando HGM es zero cuando no se encuentra una de las partes del contorno, por eso se denomina estricta. Mientras que SGM permite algo de tolerancia ya que añade un pequeño valor, $\epsilon > 0$, a cada localización, aún así la repuesta en el píxel estudiado se verá mermada. Hemos utilizado $\epsilon = 10^{-6}$, valores más pequeños producían respuestas similares a HGM. En la métrica AM, la falta de presencia de una de las partes de contorno produce un efecto pequeño en la respuesta final del filtro COSFIRE en comparación con SGM.

4.4.2 Experimentación

Hemos dividido el conjunto de imágenes en dos subconjuntos, entrenamiento y test. El conjunto de entrenamiento está formado por las imágenes HeadTool0001.bmp, HeadTool0014.bmp, HeadTool0028.bmp, HeadTool0042.bmp, HeadTool0056.bmp, HeadTool0070.bmp, HeadTool0084.bmp, HeadTool0098.bmp, HeadTool0112.bmp y HeadTool0126.bmp. Las 134 imágenes restantes conforman el conjunto de test.

Hemos configurado tantos filtros como han sido necesarios para detectar todas las plaquitas del conjunto de entrenamiento con una exhaustividad del 100 % y una precisión del 100 %. La exhaustividad es el porcentaje de plaquitas que han sido detectadas con éxito, $R = VP/(VP + FN)$ y la precisión es el ratio de plaquitas correctamente detectadas en relación a todos las respuestas positivas obtenidas $P = VP/(VP + FP)$.

El proceso iterativo comienza configurando una plaquita del conjunto de entrenamiento. Hemos configurado un filtro para el prototipo mostrado en la Fig. 9. Después hemos aplicado este filtro a todas las imágenes del conjunto de entrenamiento. Establecemos el valor de t_3 de tal modo que produzca un alto número de plaquitas correctamente detectadas sin producir ningún falso positivo, por tanto consiguien-

do una precisión del 100 %. El umbral t_3 se fijó en 0.283, 0.044 y 0.119 para AM, HGM y SGM con invarianza a rotación, detectando correctamente 9, 35 y 37 plaquitas respectivamente. En total, hay 86 plaquitas en las 10 imágenes del conjunto de entrenamiento. Por tanto, un único filtro COSFIRE detecta el 43.02 % de las plaquitas cuando se utiliza SGM.

En la segunda iteración, elegimos una de las plaquitas no detectadas por el primer filtro como nuevo prototipo. Utilizamos este prototipo para configurar un segundo filtro COSFIRE. Después, aplicamos este filtro a las 10 imágenes del conjunto de entrenamiento y asignamos nuevos valores al umbral t_3 buscando una precisión del 100 %. El nuevo filtro detecta una cantidad de plaquitas, algunas ya detectadas por el primer filtro y otras aún no detectadas. El proceso continúa sucesivamente hasta que las 86 plaquitas del conjunto de entrenamiento han sido correctamente detectadas. El número de filtros necesarios para obtener el 100 % de precisión y exhaustividad se muestra en la Tabla. 2.

Tabla 2: Resultados en términos del número de filtros COSFIRE configurados, la precisión, la exhaustividad y la media armónica para las diferentes métricas evaluadas: media artmética (AM), media geométrica extricta (HGM), media geométrica suavizada (SGM) y SGM configurando los mismos 19 filtros COSFIRE que para HGM (SGM_{19}).

	AM	**HGM**	**SGM**	**SGM_{19}**
n° filtros	24	19	17	19
Precisión (%)	81,77	**92,62**	92,25	92,39
Exhaustividad (%)	78,03	87,08	85,76	**87,52**
F-Score (%)	79,83	89,76	88,89	**89,89**

El conjunto de filtros COSFIRE así configurados se aplica al conjunto de test y se calculan los resultados en términos de precisión, exhaustividad y media armónica $F - Score$:

$$F - Score = 2 \cdot \frac{P \cdot R}{P + R} \tag{4}$$

4.4.3 Resultados

Evaluamos los resultados de la detección de plaquitas mediante un conjunto de filtros COSFIRE y comparamos los resultados utilizando las diferentes métricas presentadas, Tabla. 2. Se puede concluir que las medias geométricas son más adecuadas que las medias aritméticas para la detección de plaquitas. Para comparar las métricas SGM y HSM hemos aplicado los 19 filtros necesarios por HGM con la métrica SGM, logrando una media armónica del 89.89 % que supera a HGM en un 1.12 %.

Finalmente, cambiando los valores del parámetro t_3 se logran diferentes resultados. Un aumento en el valor de t_3 causa un incremento de la precisión y un decremento de la exhaustividad. Para cada filtro COSFIRE, sumamos (o restamos) del valor t_3 obtenido en el entrenamiento una cantidad en pasos de $0{,}01t_3$. Para todas las métricas estudiadas, la máxima media armónica se alcanza para el valor original del parámetro t_3. Los valores configurados del umbral t_3 son por tanto los que mejores resultados alcanzan en el conjunto de test.

4.4.4 Discusión

Probamos a utilizar métodos basados en características locales invariantes, pero los resultados fueron inferiores. También, se utilizaron métodos basados en correspondencia de plantillas con peores resultados (F-Score=86 %, precisión 82 % y exhaustividad 89 %) en el mismo conjunto de imágenes, (Aller-Álvarez et al., 2015). Por otra parte están los métodos basados en conocimiento previo del entorno, como el presentado en la Sección 4.2, con buenos resultados pero dependiente de la aplicación concreta a solventar. El método aquí presentado es mucho más versátil ya que puede ser aplicado para detectar cualquier herramienta o pieza sin utilizar conocimiento del entorno.

5 Reconocimiento de objetos para la recuperación de imágenes mediante ejemplos

5.1 Evaluación de configuraciones de agrupamiento de características SIFT para el reconocimiento de objetos para CBIR

5.1.1 Método

En primer lugar, obtenemos puntos clave y sus descriptores mediante SIFT (Lowe, 2004) para las regiones de interés (ROI) de los objetos de consulta y para todas las imágenes del conjunto de imágenes. Después, para cada imagen en cola, calculamos la similitud coseno entre un descriptor de la ROI con todos los descriptores de la imagen de entrada. Para este descriptor de la ROI, consideramos el emparejamiento que obtiene la máxima similitud (mínimo ángulo coseno) siempre y cuando el ángulo coseno sea al menos 2 veces el ángulo coseno del segundo vecino. De otro modo, descartamos ese emparejamiento. Repetimos este cálculo para todos los descriptores de la ROI, obteniendo un conjunto de emparejamientos entre la ROI y la imagen de entrada. Después, o bien utilizamos directamente esta información para ordenar las imágenes recuperadas o utilizamos un algoritmo de votación y una

verificación geométrica de la pose del objeto para decidir sobre la correspondencia entre imágenes.

Por un lado, consideramos la correspondencia del par de descriptores que obtuvo el ángulo coseno mínimo entre los emparejamientos de la ROI y la imagen de entrada después del test del segundo vecino más cercano. El par de puntos clave que consiguen esta correspondencia es el más similar y por tanto tiene mayor probabilidad de ser correcto. Utilizamos el valor del ángulo coseno de este par como medida de la similitud entre la ROI y la imagen de entrada. La lista de imágenes recuperadas se ordena de menor a mayor ángulo coseno. Nos referimos a este caso como *sin agrupamiento*.

Por otro lado, del conjunto de correspondencias entre los descriptores de la ROI y la imagen de entrada, identificamos conjuntos de puntos clave que votan por la misma pose de un objeto utilizando la transformada Hough y realizamos una verificación geométrica con un algoritmo de mínimos cuadrados tal y como sugiere Lowe (2004).

Cada punto clave SIFT especifica 4 parámetros: localización 2D, escala y orientación. Por tanto, podemos crear una transformada Hough que prediga el modelo de localización, orientación y escala de los puntos clave emparejados. La transformada Hough crea un acumulador de cuatro dimensiones y utiliza los parámetros de cada punto clave para votar por todas las poses del objeto que son consistentes con él mismo. Cuando conjuntos de puntos clave votan por la misma posición de un objeto, es más probable que pertenezcan al mismo objeto que si se confía únicamente en un único punto clave (Lowe, 2004). El agrupamiento de Lowe usa intervalos de tamaños amplios con 30 grados para la orientación, un factor de 2 para la escala y 0,25 veces la proyección máxima de la dimensión de la imagen de entrada para la localización (utilizando la escala a la que se definió el punto clave). Nos referimos a este caso como *agrupamiento de Lowe*.

A continuación, utilizamos un algoritmo de mínimos cuadrados para conseguir una verificación geométrica. Cada emparejamiento en un grupo debe obedecer el modelo Hough, de otro modo consideramos que esa correspondencia es un valor atípico y lo eliminamos. Si un grupo contiene menos de tres puntos después de esta eliminación de valores atípicos, rechazamos el grupo entero de correspondencias.

Finalmente para cada grupo restante, calculamos la media de los ángulos coseno de las correspondencias en ese grupo. Consideramos la menor media de todos los grupos como la medida de la similitud entre la ROI y la imagen de entrada. De nuevo, la lista de imágenes recuperadas se ordena de manera ascendente en función de esta métrica.

Evaluamos otras opciones de los parámetros utilizados en el modelo de transformada Hough. Intentamos conseguir un agrupamiento menos restrictivo de las

Tabla 3: Descripción de los objetos de consulta. Número de imágenes que contienen cada objeto de consulta en el conjunto de 614 imágenes. Tamaño en píxeles de cada ROI.

Objecto	**Número de objetos de consulta**	**Tamaño de la ROI (píxeles)**
Libro	115	305×334
Coche azul	102	285×258
Coche amarillo	138	208×265
Pinza rosa	125	146×132
Pinza azul	92	85×145
Pinza verde	42	68×59

correspondencias al aumentar el tamaño de los intervalos considerados, es decir, al disminuir el número de intervalos del acumulador. De este modo más puntos clave están de acuerdo con la misma pose del objeto. Al mismo tiempo, menos correspondencias falsas son rechazadas. Utilizamos 60 y 90 grados para la orientación, un factor de 4 y 6 para la escala y 0,5 y 0,75 veces la proyección máxima de la dimensión de la imagen de entrada para la localización en los *agrupamiento mitad* y *agrupamiento cuarto* respectivamente.

5.1.2 Evaluación

Data set. Hemos creado y publicado un conjunto de imágenes[3] para simular el contexto del proyecto ASASEC. Se compone de 614 fotogramas de 640×480 píxeles que provienen de 3 vídeos. Los vídeos fueron grabados en diferentes habitaciones y por tanto con diferente iluminación, texturas, etc. Algunos objetos están presentes en todos los vídeos como son: dos coches de juguete, algunas pinzas, una abeja de peluche, algunos bolígrafos, algunas tazas o un libro infantil junto con una muñeca. La muñeca es el actor principal en los vídeos y ayuda a simular oclusiones parciales de los objetos y un escenario más realista. Los objetos no aparecen en cada fotograma y además cada habitación tiene sus objetos propios. Junto con el conjunto de imágenes, proveemos un etiquetado indicando qué objetos se encuentran visibles en cada fotograma.

Experimentos y resultados. Como objetos de consulta hemos elegido: el libro, el coche azul y el coche amarillo y las pinzas rosa, azul y verde que mostramos en la Fig. 10. El número total de objetos de consulta presentes entre los 614 fotogramas del conjunto de imágenes y el tamaño de las ROI se especifican en la Tabla 3.

Las métricas de precisión y exhaustividad están definidas para el conjunto de imágenes completo y no miden la calidad de la ordenación de la lista de imágenes

[3]El conjunto de imágenes está disponible en http://pitia.unileon.es/varp/galleries

Libro | Coche azul | Coche amarillo

Pinza rosa | Pinza azul | Pinza verde

Figura 10: ROI de los objetos de consulta.

recuperadas. La relevancia del ranking puede medirse calculando la precisión a diferentes puntos de corte, lo que se conoce como precisión en n o P@n. Denominamos a h $[i]$ la imagen recuperada en la posición i en la lista de recuperación, rel $[i]$ es 1 si h $[i]$ es relevante y 0 de otro modo. Para que una imagen recuperada sea relevante el objeto de interés tiene que estar presente en la imagen y, además, correctamente localizado. De este modo, la precisión en n es:

$$\text{P@}n = \sum_{k=1..n} \text{rel}\,[k]\,/n \tag{5}$$

La Tabla 4 muestra los resultados para los cuatro tipos de agrupamientos con las precisiones para diferentes cortes.

En vista a los resultados presentados en la Tabla 4 no utilizar agrupamiento es más conveniente para conseguir altas precisiones en cortes pequeños de la lista de imágenes recuperadas, sin embargo, para conseguir altas precisiones a cortes más altos el agrupamiento propuesto por Lowe es más adecuado.

Algunos de los objetos mal clasificados provienen de confusiones con objetos de forma similar y diferentes colores ya que el método SIFT no es invariante a color. Esto podría ser solventado utilizando la versión de SIFT para imágenes en color presentada en (Van de Sande et al., 2010). Por otra parte, estamos utilizando como

Tabla 4: Precisión en diferentes cortes en función de los objetos de consulta usando diferentes agrupamientos. Los mejores resultados para cada precisión en n aparecen en negrita. Sin agrup = *sin agrupamiento*, cuarto = *agrupamiento cuarto*, mitad = *agrupamiento mitad*, Lowe = *agrupamiento Lowe*

	Libro					Pinza azul		
	P@40	P@50	P@60	P@70	P@80	P@5	P@10	P@20
Sin agrup.	**1**	**1**	**0.9**	0.8	0.75	**1**	**0.7**	**0.35**
Cuarto	0.85	0.82	0.77	0.7	0.66	0.4	0.2	0.1
Mitad	0.93	0.88	0.88	**0.83**	**0.8**	0.4	0.2	0.1
Lowe	**1**	0.96	0.85	**0.83**	0.79	0.8	0.4	0.2
	Coche azul					Pinza rosa		
	P@5	P@10	P@20	P@30	P@40	P@5	P@10	P@20
Sin agrup.	**1**	**1**	0.75	0.57	0.43	**0.8**	**0.4**	**0.25**
Cuarto	1	0.8	0.6	0.43	0.38	0.2	0.1	0.05
Mitad	0.8	0.9	0.85	0.7	0.55	0.2	0.2	0.1
Lowe	0.8	0.9	**0.9**	**0.73**	**0.625**	**0.8**	**0.4**	**0.25**
	Coche amarillo					Pinza verde		
	P@5	P@10	P@20	P@30	P@40	P@5	P@10	P@20
Sin agrup.	**1**	0.9	0.75	**0.73**	0.63	**1**	0.5	0.3
Cuarto	0.8	0.7	0.55	0.47	0.38	0.2	0.1	0.05
Mitad	**1**	0.8	0.75	0.7	**0.68**	0.8	0.4	0.2
Lowe	**1**	**1**	**0.85**	0.7	0.65	**1**	**0.7**	**0.35**

objeto de búsqueda, una región rectangular que incluye el objeto y el fondo. Los fondos de imagen muy dibujados, como por ejemplo el edredón estampado, producen puntos clave muy característicos. Algunas correspondencias erróneas se deben a emparejamientos entre puntos clave de fondos de imágenes. Un método para la segmentación automática del objeto sería de utilidad.

5.2 Añadiendo descripción de color a los filtros COSFIRE

5.2.1 Método con aplicación para la localización de vértices de color

Configuración de un filtro de color COSFIRE para la localización de vértices. Construimos el filtro de color COSFIRE a partir de las respuestas de filtros Gabor 2D aplicados a cada canal de color de una imagen. Filtrar individualmente los tres canales de color y después combinar estas tres respuestas incrementa la invarianza a iluminación y el poder de discriminación, lo que conduce a una detección más precisa de las activaciones de la imagen que filtrar únicamente el canal de luminancia (Van de Sande et al., 2010), como es la imagen en escala de grises.

Denotamos como $g_{\lambda,\theta,\zeta,c}(x,y)$ la respuesta de un filtro Gabor de longitud de onda λ, orientación θ y desplazamiento de fase ζ en un canal de color dado c de la imagen prototipo P. Los filtros Gabor pueden ser simétricos ($\zeta \in \{0, \pi\}$), asimétricos ($\zeta \in \left\{\frac{\pi}{2}, \frac{3\pi}{2}\right\}$) o un filtro de energía, para más detalles referimos al lector a (Petkov, 1995; Petkov and Kruizinga, 1997; Kruizinga and Petkov, 1999; Grigorescu et al., 2002; Petkov and Westenberg, 2003; Grigorescu et al., 2003b,a). Normalizamos las funciones Gabor para que todas los valores positivos sumen 1 y los negativos sumen -1.

La respuesta de un filtro Gabor se obtiene al convolucionar la imagen de entrada con un kernel Gabor. Nosotros obtenemos un nuevo kernel por cada kernel Gabor utilizado. Para filtros simétricos, el nuevo kernel se compone de la parte central del kernel Gabor mientras que para filtros asimétricos se compone de la parte positiva más larga del kernel Gabor. Denotamos mediante $K_{\lambda,\theta,\zeta}$ dichos kerneles asociados con sus correspondientes respuestas Gabor $g_{\lambda,\theta,\zeta,c}(x,y)$.

Calculamos la norma L-infinito de las tres respuestas Gabor obtenidas para cada canal de color.

$$g_{\lambda,\theta,\zeta}(x,y) = \max_{z=1,2,3} \{g_{\lambda,\theta,\zeta,c_z}(x,y)\} \tag{6}$$

A continuación, calculamos la norma L-infinito para los valores de ζ empleados. Utilizamos $\zeta = \{0, \pi\}$ para la detección de líneas y $\zeta = \left\{\frac{\pi}{2}, \frac{3\pi}{2}\right\}$ para la detección de bordes. Analizamos ambos valores de ζ para conseguir independencia de la luminancia del fondo de la imagen.

$$g_{\lambda,\theta}(x,y) = \max_{z=1,2} \{g_{\lambda,\theta,\zeta_z}(x,y)\} \tag{7}$$

Finalmente, umbralizamos las respuestas de los filtros Gabor para fracciones t_1 ($0 \leq t_1 \leq 1$) de la máxima respuesta de $g_{\lambda,\theta}(x,y)$ para todas las combinaciones de valores (λ,θ) utilizados y todas las posiciones (x,y) en la imagen, y denotemos estas respuestas umbralizadas como $|g_{\lambda,\theta}(x,y)|_{t_1}$.

El filtro de color COSFIRE se configura alrededor de un punto de interés, que consideramos el centro del filtro. Tomamos las respuestas de un banco de filtros Gabor, caracterizadas por los parámetros (λ,θ), a lo largo de circunferencias de radios dados ρ alrededor de un punto de interés. El filtro de color COSFIRE está definido en ciertas posiciones (ρ_i, ϕ_i) con respecto al punto de interés en las cuales se producen máximos locales de las respuestas del banco de filtros Gabor. Un conjunto de siete parámetros $(\lambda_i, \theta_i, \rho_i, \phi_i, \gamma_{1_i}, \gamma_{2_i}, \gamma_{3_i})$ caracteriza las propiedades de una parte de contorno que está presente en el patrón de interés: $\lambda_i/2$ representa la anchura, θ_i representa la orientación, (ρ_i, ϕ_i) representa la localización y $(\gamma_{1_i}, \gamma_{2_i}, \gamma_{3_i})$ representa la descripción de color de cada canal.

La obtención de los parámetros $(\lambda_i, \theta_i, \rho_i, \phi_i)$ se realiza siguiendo a Azzopardi and Petkov (2013c). Para obtener la descripción de color de las tuplas, calculamos la media de los valores de los píxeles en una región alrededor de la localización (ρ_i, ϕ_i) para cada canal de color. Centramos el kernel $K_{\lambda_i,\theta_i,\zeta}$ en la posición (ρ_i, ϕ_i) y realizamos una multiplicación píxel a píxel del kernel y el canal de color de la imagen prototipo P_c. Después normalizamos el resultado. De este modo obtenemos un valor que describe el color de cada canal de color en la localización considerada, γ_{c_i}.

$$\gamma_{c_i} = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n} P_c(x_i + k - 1, y_i + l - 1) K_{\lambda_i,\theta_i,\zeta}(k,l)}{\sum_{k=1}^{m}\sum_{l=1}^{n} K_{\lambda_i,\theta_i,\zeta}(k,l)} \tag{8}$$

donde m y n son las filas y columnas del kernel $K_{\lambda_i,\theta_i,\zeta}$ respectivamente y (x_i, y_i) las coordenadas cartesianas de (ρ_i, ϕ_i). Se calcula la media en lugar de tomar el valor del píxel directamente para evitar que posibles píxeles que presenten ruido puedan afectar profundamente la descripción de color. En filtros Gabor simétricos, los kernel son idénticos independientemente del valor de ζ, cualquiera puede ser utilizado. Además, como utilizamos la parte central del kernel Gabor normalizado, aseguramos que la descripción de color se calcula en una región cuyo ancho es como máximo, $\lambda_i/2$, el ancho de la línea localizada por el método. En filtros Gabor antisimétricos, usamos el kernel $K_{\lambda_i,\theta_i,\zeta}$ con el valor de $\zeta \in \left\{\frac{\pi}{2}, \frac{3\pi}{2}\right\}$ para el que la distancia Euclídea desde el centroide del kernel al punto de interés sea mínima cuando ambos kernel están centrados alrededor de la localización (ρ_i, ϕ_i). De este modo, describimos la parte del prototipo más cercana al centro del filtro color COSFIRE.

El conjunto $S_f = \{p_i | i = 1 \dots n_c\} = \{(\lambda_i,\theta_i,\rho_i,\phi_i,\gamma_{1_i},\gamma_{2_i},\gamma_{3_i}) \mid i = 1 \dots n_c\}$ denota las combinaciones de parámetros que cumplen las condiciones previas. El subíndice f hace referencia al patrón del prototipo alrededor del punto de interés seleccionado y n_c es el número de partes de contorno localizadas.

Aplicación de un filtro de color COSFIRE para la localización de vértices. Calculamos la respuesta de un banco de filtros Gabor 2D aplicados a cada canal de color de la imagen de entrada para los pares de valores (λ_i, θ_i) del conjunto S_f y para ambos valores de ζ. Ambos valores de ζ se analizan para que el método localice el patrón de interés independientemente del fondo de la imagen. Después, calculamos dos normas L-infinito consecutivas, a lo largo de los canales de color y de los valores de ζ. A continuación, umbralizamos las respuestas a una fracción t_1 de la máxima respuesta, resultando en una respuesta Gabor $|g_{\lambda_i,\theta_i}(x,y)|_{t_1}$ para cada tupla p_i en el conjunto S_f. También obtenemos los kernel $K_{\lambda_i,\theta_i,\zeta}$ asociados con los filtro Gabor para cada tupla.

Las respuestas de los filtros Gabor son emborronadas para permitir tolerancia en

las posiciones de las partes del contorno. Definimos el emborronado como una operación de convolución de las repuestas Gabor umbralizadas $|g_{\lambda_i,\theta_i}(x,y)|_{t_1}$ con un filtro paso bajo Gaussiano invariante a rotación $G_\sigma(x,y)$ de tamaño $1 \times n\sigma$ píxeles con desviación estándar σ. La desviación estándar es una función lineal de la distancia ρ como en (Azzopardi and Petkov, 2013c). La respuesta emborronada para una tupla p_i es:

$$b_{\lambda_i,\theta_i,\rho_i}(x,y) = |g_{\lambda_i,\theta_i}(x,y)|_{t_1} * G_{\sigma_i}(x,y) * G'_{\sigma_i}(x,y) \tag{9}$$

Después, desplazamos las respuestas emborronadas de cada tupla p_i una distancia de ρ_i en la dirección opuesta a ϕ_i. Denotamos por $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ las respuestas emborronadas y desplazadas del filtro de Gabor especificado en la tupla p_i del conjunto S_f.

Por otro lado, calculamos las respuestas a las descripciones de color. Para cada tupla, p_i, convolucionamos cada canal de color de una imagen de entrada I_c con los correspondientes kernel deslizantes $K_{\lambda_i,\theta_i,\zeta}$ y después normalizamos los resultados.

$$v_{\lambda_i,\theta_i,c}(x,y) = \frac{I_c(x,y) * K_{\lambda_i,\theta_i,\zeta}(x,y)}{\sum_{k=1}^{m}\sum_{l=1}^{n} K_{\lambda_i,\theta_i,\zeta}(k,l)} \tag{10}$$

donde k y l son las filas y columnas del kernel $K_{\lambda_i,\theta_i,\zeta}$ respectivamente.

Para filtros Gabor antisimétricos, de nuevo, calculamos las convoluciones para los dos valores de ζ.

Llamamos $h_{p_i}(x,y)$ a la respuesta de la descripción de color de la tupla p_i en el conjunto S_f. Calculamos $h_{p_i}(x,y)$ aplicando un kernel Gaussiano que mide la similitud entre los colores de la parte de contorno del prototipo y los colores de la imagen de entrada en cada canal de color siguiendo la Eq. 11.

$$h_{p_i}(x,y) = exp^{-\frac{\sum_{c=1}^{3}[v_{\lambda_i,\theta_i,c}(x,y)-\gamma_{c_i}]^2}{2\sigma_g^2}} \tag{11}$$

donde σ_g es la desviación estándar del kernel de color Gaussiano.

Para filtros Gabor antisimétricos, calculamos una respuesta de descripción de color para cada kernel Gabor y después obtenemos el máximo valor de las dos respuestas para cada par de píxeles correspondientes (x_j, y_j).

$$h'_{p_i}(x,y) = \max_{j,k}\{h_{p_i}(x_j,y_k)|\zeta = \pi/2,\ h_{p_i}(x_j,y_k)|\zeta = 3\pi/2\} \tag{12}$$

Después, emborronamos la respuesta de descripción de color, $h''_{p_i}(x,y)$. Finalmente, desplazamos la respuesta emborronada una distancia ρ_i en el sentido contrario a ϕ_i, $\hat{h}_{p_i}(x,y)$.

Definimos la respuesta a un filtro de color COSFIRE $r_{S_f}(x,y)$ como la media

geométrica ponderada del producto Hadamard de las respuestas de los filtros Gabor emborronadas y desplazadas, $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$, por las respuestas de color Gaussianas emborronadas y desplazadas, $\hat{h}_{p_i}(x,y)$, que corresponden a las propiedades de las partes de contorno descritas en S_f:

$$r_{S_f}(x,y) \stackrel{\text{def}}{=\!=} \left(\prod_{i=1}^{|S_f|} \left(s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \circ \hat{h}_{p_i}(x,y) \right)^{\omega_i} \right)^{1/\sum_{i=1}^{|S_f|}\omega_i} \tag{13}$$

$$\omega_i = exp^{-\frac{\rho_i^2}{2\sigma'^2}} \tag{14}$$

$$\sigma' = \left(-\rho_{max}^2/2ln\tau\right)^{1/2} \tag{15}$$

$$\rho_{max} = \max_{i \in 1 \ldots |S_f|} \{\rho_i\} \tag{16}$$

donde $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y) \circ \hat{h}_{p_i}(x,y)$ indica el producto Hadamard de $s_{\lambda_i,\theta_i,\rho_i,\phi_i}(x,y)$ y $\hat{h}_{p_i}(x,y)$.

Finalmente, umbralizamos la respuesta del filtro de color COSFIRE a una fracción t_3 del máximo global para cada coordenada de la imagen (x,y), $0 \leq t_3 \leq 1$.

$$r(x,y) = \left| r_{S_f}(x,y) \right|_{t_3} \tag{17}$$

5.2.2 Método con aplicación para la localización de objetos de color

Configuración de un filtro de color COSFIRE para la localización de objetos. Utilizamos el detector SIFT (Lowe, 2004) para buscar puntos clave estables en el prototipo. Del punto clave SIFT, estamos interesados en sus coordenadas y la escala. Aplicamos el detector SIFT a cada canal de la imagen de entrada I_c y consideramos los puntos clave cuya escala es mayor que una fracción t_4 de la máxima escala de los puntos clave. Después agrupamos los puntos clave restantes en tres grupos de acuerdo con sus valores de escala utilizando el algoritmo k-means (Duda et al., 2000), y asignando a cada punto clave el valor de escala medio del grupo al que pertenece. Este paso no es esencial pero permite acelerar el cálculo. Finalmente, sólo los puntos clave únicos se conservan, (δ_j, x_j, y_j).

El punto de interés del prototipo (x_p, y_p), que es el centro del filtro de color COSFIRE, puede ser elegido manualmente o asignado automáticamente al centro de la ROI. Calculamos las coordenadas locales (ρ_j, ϕ_j) de los puntos clave (x_j, y_j) con respecto al punto de interés del patrón prototipo.

$$(\rho_j, \phi_j) = \left(\sqrt{(x_j - x_p)^2, (y_j - y_p)^2}, \text{ atan2}(y_j - y_p, x_j - x_p) \right) \tag{18}$$

donde atan2 es el ángulo en radianes entre el eje x positivo de un plano y el punto dado por las coordenadas (x_j, y_j) en él.

Para cada punto clave $(\delta_j, \rho_j, \phi_j)$, creamos una máscara circular Gaussiana $K_{\delta_j,\rho_j,\phi_j}(x, y)$ de radio δ_j centrada en las correspondientes posiciones (ρ_j, ϕ_j).

$$K_{\delta_j,\rho_j,\phi_j}(x, y) = \exp^{-\frac{x^2+y^2}{2(\delta/2)^2}} \tag{19}$$

Después realizamos una multiplicación píxel a píxel de la máscara por cada canal de color del prototipo P_c y normalizamos los resultados. Por tanto, obtenemos un valor de descripción de color para cada canal de color γ_{c_j} en el punto clave considerado $(\delta_j, \rho_j, \phi_j)$.

$$\gamma_{c_j} = \frac{\sum_{k=1}^{m} \sum_{l=1}^{n} P_c(x_j + k - 1, y_j + l - 1) K_{\delta,x_j,y_j}(k, l)}{\sum_{k=1}^{m} \sum_{l=1}^{n} K_{\delta,x_j,y_j}(k, l)} \tag{20}$$

donde m y n son las filas y columnas del kernel K_{δ,x_j,y_j} respectivamente y (xj, y_j) las coordenadas cartesianas de (ρ_j, ϕ_j).

Un conjunto de seis parámetros o tupla $p_j = (\delta_j, \rho_j, \phi_j, \gamma_{1_j}, \gamma_{2_j}, \gamma_{3_j})$ especifica las propiedades de una parte de contorno en este nuevo conjunto $S'_f = \{p_j | j = 1 \dots n_k\} = \{(\delta_j,\rho_j,\phi_j,\gamma_{1_j},\gamma_{2_j},\gamma_{3_j}) \mid j = 1 \dots n_k\}$. Donde n_k hace referencia al número de puntos clave detectados.

Calculamos otro conjunto de tuplas $S_f = \{p_i | i = 1 \dots n_c\} = \{(\lambda_i,\theta_i,\rho_i,\phi_i,\gamma_{1_i},\gamma_{2_i}, \gamma_{3_i}) \mid i = 1 \dots n_c\}$ para el objeto de interés, tal y como se explicó en la Sección 5.2.1 usando un banco de filtros Gabor antisimétricos.

Aplicación de un filtro de color COSFIRE para la localización de objetos. Para cada valor único de δ_j en las tuplas de S'_f, calculamos una máscara circular Gaussiana $K_{\delta_j}(x, y)$ que contiene un círculo de radio δ_j. Después convolucionamos cada canal de color de la imagen de entrada I_c con la máscara $K_{\delta_j}(x, y)$ y normalizamos los resultados.

Denominamos $d_{p_j}(x, y)$ la respuesta de descripción de color de los puntos clave para la tupla p_j en el conjunto S'_f. Calculamos $d_{p_j}(x, y)$ aplicando un kernel Gaussiano que mida la similitud entre los colores de la parte de contorno definida por la tupla p_j y los colores de la imagen de entrada normalizada y convolucionada correspondientemente para cada canal de color.

Después, emborronamos la respuesta y la desplazamos una distancia ρ_j en la dirección contraria a ϕ_j, obteniendo $\hat{d}_{p_j}$.

Definimos la respuesta $r_{S'_f}(x, y)$ de un filtro de color COSFIRE para la descrip-

ción de puntos clave en un objeto de interés como la media geométrica ponderada de las respuestas de similitud Gaussianas emborronadas y desplazadas $d_{p_j}(x,y)$ que corresponden con las propiedades de las partes de contorno descritas en S'_f:

$$r_{S'_f}(x,y) \stackrel{\text{def}}{=\!=} \left(\prod_{j=1}^{|S'_f|} \left(\hat{d}_{p_j}(x,y) \right)^{\omega_j} \right)^{1/\sum_{j=1}^{|S'_f|} \omega_j} \tag{21}$$

donde ω_j se define en Eq. 14.

Calculamos la respuesta de un filtro de color COSFIRE $r(x,y)$ como el producto Hadamard umbralizado de las respuestas de detección de bordes de color y descripción de puntos clave de color:

$$r(x,y) \stackrel{\text{def}}{=\!=} \left| r_{S_f}(x,y) \circ r_{S'_f}(x,y) \right|_{t_5} \tag{22}$$

donde $||_{t_5}$ significa que la respuesta es umbralizada a una fracción t_5 de su máximo en las coordenadas (x,y).

La Fig. 11 muestra la aplicación de un filtro de color COSFIRE para la localización de objetos en color. La respuesta del filtro de color COSFIRE es el producto Hadamard de la media geométrica ponderada de 12 respuestas de la descripción a color de puntos clave y la media geométrica ponderada de 67 respuestas para la detección de bordes de color. El filtro responde en puntos donde hay un patrón idéntico o similar al objeto prototipo de interés y en el punto de interés de dicho objeto a pesar de los diferentes colores y patrones del fondo del objeto.

5.2.3 Experimentos y resultados

Utilizamos el conjunto de imágenes público COIL-100 para realizar la experimentación. Configuramos un filtro color COSFIRE para cada clase, en concreto, para la imagen con un ángulo de rotación de 0°. También configuramos un filtro COSFIRE estándar para las mismas imágenes de modo que los parámetros comunes fueran iguales. Aplicamos cada filtro COSFIRE a todo el conjunto de imágenes y calculamos la precisión y exhaustividad para cada posición de la lista de imágenes recuperadas. Calculamos la precisión promedio, AveP, que es el área bajo la curva precisión-exhaustividad y la media de la precisión promedio, MAP, para el conjunto de consultas como el promedio de las puntuaciones medias de precisión para cada consulta. A su vez, calculamos el promedio de las medias armónicas máximas de precisión y exhaustividad para todas las consultas del conjunto de imágenes, MFScore. La precisión media, MPrecision, y la exhaustividad media, MRecall, son los promedios de las precisiones y exhaustividades, respectivamente, que obtuvieron

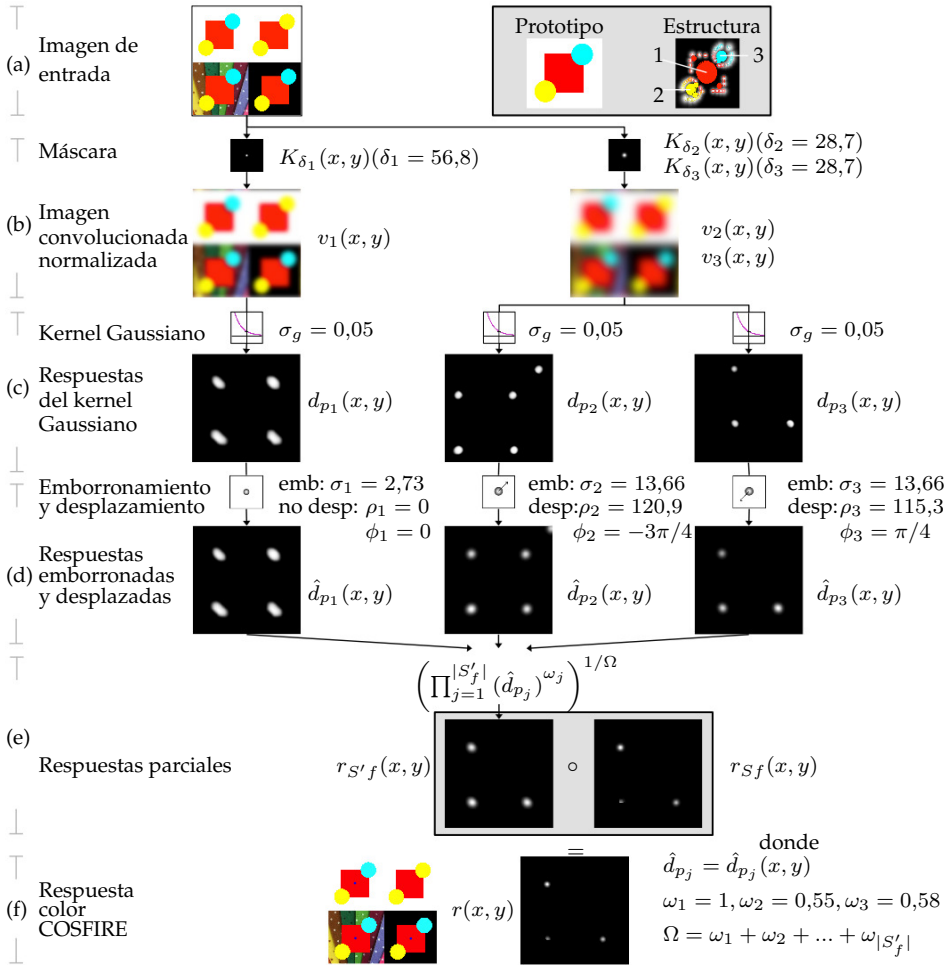


Figura 11: Aplicación de los filtros color COSFIRE. (a) Imagen de entrada, prototipo y estructura del filtro configurado. Los números indican tres tuplas en S'_f para los que ilustramos esta aplicación. (b) Convolución normalizada de la imagen de entrada con una máscara circular Gaussiana. (c) Similitud entre los colores de las partes de contorno y los colores de la imagen de entrada medida con un kernel Gaussiano. (d) Emborronamos y desplazamos las respuestas previas hacia el centro del prototipo. (e) Respuesta para la descripción de color de los puntos clave. (f) Respuesta del filtro de color COSFIRE. Hay tres máximos locales que se corresponden con los tres objetos similares al de interés en la imagen de entrada. emb=emborronamiento, desp=desplazamiento.

medias armónicas máximas para todas las consultas del conjunto de imágenes. Los resultados se muestran en la Tabla 5, donde se observa la efectividad de los filtros

color COSFIRE frente a los filtros COSFIRE tradicionales.

Tabla 5: Media de la precisión promedio, MAP; promedio de las medias armónicas, MFScore; precisión media, MPrecision; y exhaustividad media, MRecall, del conjunto de imágenes COIL para los filtros color COSFIRE, C, y los filtros COSFIRE tradicionales, G.

	C	**G**
MAP	0.6970	0.1322
MFScore	0.7617	0.2241
MPrecision	0.9388	0.3217
MRecall	0.6822	0.3162

También evaluamos los resultados de los filtros COSFIRE ante un problema de clasificación. Las respuestas de un fitro COSFIRE dado se dividen por la máxima respuesta obtenida por dicho filtro. Una imagen dada se clasifica como perteneciente a la clase para la que el filtro COSFIRE que obtuvo la máxima respuesta fue configurado. Calculamos una matriz de confusión donde cada posición (i, j) es el número de imágenes de la clase i clasificadas como clase j. Las Fig. 12 and 13 muestran las matrices de confusión de los filtros color COSFIRE y COSFIRE tradicional, respectivamente. La matriz de confusión de los filtros color COSFIRE es menos dispersa que la del método tradicional, y presenta altos valores en la diagonal y valores bajos fuera de la diagonal. El método propuesto basado en color obtiene una tasa de acierto del 67.57 % mientras que el método tradicional sólo consigue el 21.69 %, calculando la tasa de acierto como la traza de la matriz de confusión dividido entre el número total de imágenes del conjunto.

6 Conclusiones

Las contribuciones del trabajo presentado en esta tesis ayudan al entendimiento y la resolución de aplicaciones reales utilizando técnicas de reconocimiento de objetos y clasificación de imágenes.

Algunas conclusiones específicas que podemos extraer de este trabajo son:

1. Los métodos basados en ILF nunca habían sido utilizados en la evaluación de la integridad del acrosoma. Demostramos el éxito de aplicarlos para la evaluación del estado de los acrosomas de verraco como intactos y dañados. SURF obtuvo una tasa de acierto del 94.88 % con k-NN, mejorando los descriptores de textura global y los trabajos previos en la fecha en la que los resultados fueron publicados como un artículo de conferencia. Además, se observó que SURF y SIFT obtuvieron mejores resultados para la clase de dañados que para

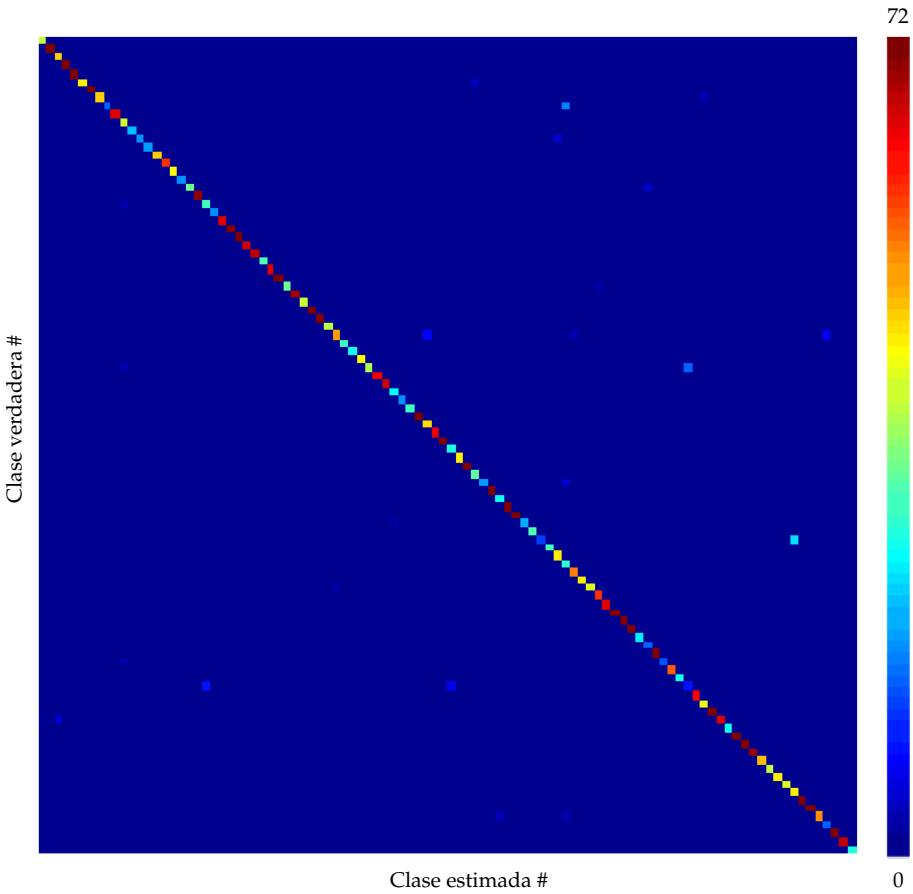


Figura 12: Matriz de confusión para los filtros color COSFIRE propuestos. La matriz tiene un de tamaño 100×100. Las columnas representan el número de predicciones de cada clase y las filas las instancias en la clase real.

la de intactos mientras que los descriptores de textura globales se comportan de forma contraria.

2. En la misma línea de trabajo, propusimos un enfoque para clasificar las características SURF con SVM, sin utilizar BoW. La clasificación de cabezas obtuvo una tasa de acierto del 90.91 % mejorando la clasificación de simples puntos clave. Este enfoque puede ser fácilmente implementado para otros métodos de ILF y otros algoritmos de clasificación.

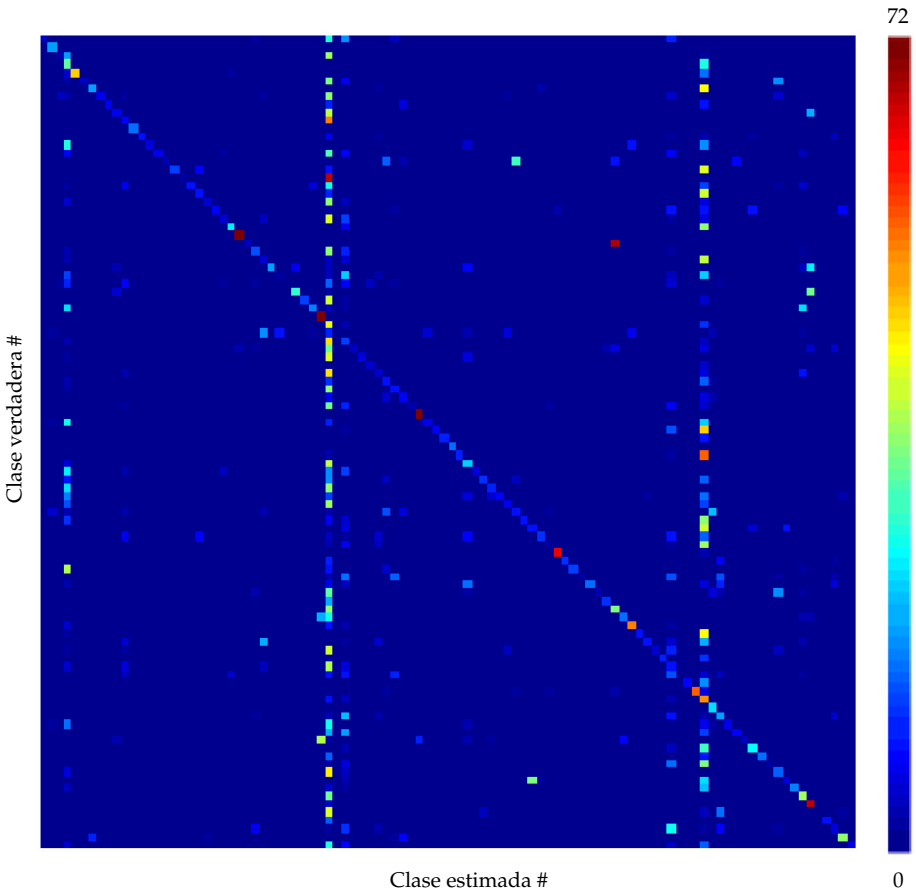


Figura 13: Matriz de confusión para los filtros COSFIRE tradicionales. La matriz tiene un de tamaño 100×100. Las columnas representan el número de predicciones de cada clase y las filas las instancias en la clase real.

3. Los resultados de la fusión temprana propuesta de ILF con descriptores globales de textura para la clasificación de la integridad de los acrosomas mejoró los métodos individuales. La concatenación de SURF con Legendre consiguió una tasa de acierto del 95.56 % con k-NN. Este es un resultado satisfactorio de acuerdo con la comunidad de veterinarios.

4. Se presentó un método altamente efectivo y eficiente para la localización de filos de corte en fresadoras. La salida es un conjunto de regiones alrededor de

los filos de corte, que pueden ser utilizadas como entradas para otros métodos que realicen la evaluación de la calidad de los bordes. Se basa en la transformada circular Hough para encontrar los tornillos que aprietan las plaquitas y en detección de bordes y la transformada de Hough estándar para localizar el filo de corte. Obtuvo un precisión del 99.61 %, definiendo precisión como la media de las fracciones de los filos de corte reales que se encuentran dentro de las ROI localizadas por el método de 20 píxeles de ancho en imágenes de 1280×960 píxeles.

5. Se introdujo un método nuevo para la descripción efectiva y la clasificación de las plaquitas con respecto al estado de los filos de corte como rotas y no rotas. Calcula las magnitudes gradiente y las desviaciones a lo largo de los filos de corte. El tiempo requerido por este método para la inspección de un cabezal de fresado es inferior al tiempo de reposo de la herramienta. Obtuvimos una media armónica de $0{,}9143(\pm 0{,}079)$ con una precisión del $0{,}9661(\pm 0{,}073)$ y una exhaustividad del $0{,}8821(\pm 0{,}134)$ en un conjunto de imágenes público con 180 plaquitas cuando calculamos resultados medios en 20 conjuntos de validación aleatorios.

6. Se ha presentado otro método para la localización de plaquitas. Es más general que el anterior ya que considera de manera independiente cada imagen del conjunto de imágenes. Está basado en filtros COSFIRE y puede ser automáticamente configurado a pesar de la apariencia de las plaquitas. Una nueva métrica, media geométrica suave, para el cálculo de la respuesta del filtro COSFIRE fue introducida, mejorando las anteriores. Esta métrica añade un pequeño valor a todas las respuestas de filtros Gabor después de su procesamiento, de este modo se provee de tolerancia a partes del contorno no localizadas. Obtuvo una media armónica de 89.89 % con precisión igual a 92.39 % y exhaustividad del 87.52 %, mejorando resultados previos basados en correspondencia de plantillas.

7. Evaluamos diferentes configuraciones de agrupamiento de puntos clave SIFT en relación con los parámetros de pose: localización de las coordenadas, escala y orientación. Precisiones más altas fueron obtenidas sin agrupamiento para pequeños cortes de la lista de imágenes recuperadas mientras que mejores resultados se obtuvieron con el agrupamiento propuesto por Lowe en cortes más altos. Los resultados fueron calculados en un conjunto de 614 imágenes que ilustran posibles escenarios de la base de datos del proyecto ASASEC.

8. Se propusieron los filtros color COSFIRE. Añaden descripción de color y poder de discriminación a los filtros COSFIRE a la vez que proveen de invarianza a la

intensidad del fondo del objeto. Los filtros color COSFIRE fueron presentados para patrones que consisten tanto en líneas de color como en objetos de color.

En el futuro se planea trabajar en varias líneas:

1. Mejora en la captura de las imágenes del cabezal de fresado para que los bordes estén más marcados. Podrían realizarse pruebas con diferentes iluminaciones o incluso capturando varias imágenes con diversas iluminaciones.

2. Implementación de un método de selección de las partes de contorno de los filtros COSFIRE más significativas. Pruebas preliminares han demostrado unas mejores tasas de éxito y mayor rapidez en la aplicación de localización de plaquitas. Este enfoque se pretende extender a otras tareas de reconocimiento de objetos.

3. Aplicación de los filtros color COSFIRE en más conjuntos de imágenes para demostrar su poder.

4. Desarrollo de un enfoque COSFIRE que se base en combinación de respuestas de puntos clave color SIFT (Van de Sande et al., 2010) en lugar de filtros Gabor. Resultados previos utilizando SIFT con COSFIRE en imágenes en escala de grises parecen prometedores.