UNIVERSITY OF LEÓN

DEPARTMENT OF ELECTRICAL, SYSTEMS AND AUTOMATIC ENGINEERING

METHODS FOR IMPROVING TEXTURE DESCRIPTION BY USING STATISTICAL INFORMATION EXTRACTED FROM THE IMAGE GRADIENT

A dissertation supervised by promotors

DR. ENRIQUE ALEGRE GUTIÉRREZ AND

DR. LAURA FERNÁNDEZ ROBLES

and submitted by

OSCAR GARCÍA-OLALLA OLIVERA

in fulfillment of the requirements for the Degree of

PHILOSOPHIÆDOCTOR (PH.D.)

León, July 2017

UNIVERSIDAD DE LEÓN

DEPARTAMENTO DE INGENIERÍA ELÉCTRICA Y DE SISTEMAS Y AUTOMÁTICA

TÉCNICAS PARA LA MEJORA DE LA DESCRIPCIÓN DE TEXTURA UTILIZANDO INFORMACIÓN ESTADÍSTICA DEL GRADIENTE DE LA IMAGEN

Tesis doctoral dirigida por

EL DR. ENRIQUE ALEGRE GUTIÉRREZ Y

LA DRA. LAURA FERNÁNDEZ ROBLES

y desarrollada por

OSCAR GARCÍA-OLALLA OLIVERA

a fin de optar al grado de

DOCTOR POR LA UNIVERSIDAD DE LEÓN

León, Julio de 2017

Abstract

This thesis proposes three new descriptors developed to the same end of taking advantage of statistical information extracted from the image gradient to improve the description of the texture present in an image.

First, we proposed a new method, Adaptive Local Binary Patterns with oriented Standard deviation (ALBPS), based on the lack of information about the orientation in the well-known Local Binary Patterns (LBP). This idea was introduced previously in Adaptive Local Binary Patterns (ALBP), that takes into account the oriented mean and standard deviation of the local absolute difference in order to make the matching more robust against local spatial structure changes. Our proposal included the standard deviation information of the image gradient along different orientations not in the matching method but in the descriptor algorithm and it is called ALBPS on that account.

Given the good results obtained by using ALBPS in comparison with LBP and most of its variants, ALBP included, we developed another new descriptor based on ALBPS that we named LOSIB (Local Oriented Statistical Information Booster). LOSIB is designed to improve the performance of other texture descriptors that mainly rely on local characteristics.It enhances description by extracting gray level differences along several orientations. Specifically, the mean of the differences along particular orientations is considered. Two additional parameters, i.e. the radius of the neighbourhood and the number of neighbours, give LOSIB more reliability and robustness than ALBPS, adapting the fetures extraction to different kinds of images and problems.

Finally, we proposed CLOSIB (Complete LOSIB). It is a generalized version of LOSIB that incorporates three parameters into its configuration, i.e. the statistical order of the descriptor, the radius of the neighbourhood and the number of neighbours. CLOSIB is presented through a comprehensive mathematical expression and includes three different variants to provide a faster alternative, H-CLOSIB, a multiscale descriptor, M-CLOSIB, and a combination of both, HM-CLOSIB.

The main common feature of the proposed methods is that they retrieve global statistical information of the texture of the image in contrast to the local descrip-

tion provided by their predecessor LBP. We also used our methods in combination with other state-of-the-art local descriptors and the early fusion of both resulted in a higher discrimative power of the final descriptor over many experimental fields tested. Particularly, the concatenation of CLOSIB with methods based on Local Binary Patterns (LBP) works fine because the characteristics used in their feature extraction are complementary.

ALBPS was evaluated for material recognition and also for vitality assessment of boar spermatozoas. Regarding the first application, ALBPS achived a 61.47% of hit rate on KTH Tips2-a dataset outperforming all the other LBP-based methods evaluated. In relation to the latter field, ALBPS combined with 13 Wavelet Concurrent Features (WCF13) global descriptor, that extracts Haralick features from the Wavelet Transform, yielded up an F-Score of 0.886. This is the best result achieved so far in this field up to our knowledge.

Three diverse application fields were evaluated to test the efficiency of LOSIB, which are materials recognition, classification of the acrosome integrity of boar spermatozoa and tool wear estimation. For materials recognition, two different datasets were evaluated: KTH-Tips-2a and Brodatz32 to prove the robustness of LOSIB. LOSIB was used together with LBP and several of its state-of-the-art variants. Combined with CLBP (Complete LBP), results were improved in 5.80% on KTH-Tips 2a and 7.09% on the Brodatz dataset. All tested descriptors in combination with LOSIB achieved a higher performance on both datasets using two different classifiers. Regarding acrosome integrity assessment of boar spermatozoa, three descriptors were combined: 13 Haralick features on the Gray Level Co-oncurrent Matrix (GLCM) of the Haar DWT; a texture description using LBP and a shape description based on Fourier. This method, called WFLP, was combined with LOSIB obtaining a 100% of accuracy and a 0.9975 of FScore, being able to precisely determine the state of the boar acrosomes without using neither stains nor expensive microscopes. For tool wear estimation, LOSIB combined wih ALBP achieved a 74.06% of hit rate for a binary classification of the wear level of milling inserts.

CLOSIB was evaluated for three applications, i.e. material recognition –by means of KTH TIPS (2-a) dataset–, face recognition –using JAFFE dataset– and textile recognition –by means of creating and introducing a new dataset for textile retrieval in indoor rooms under challenging conditions of shape, illumination and occlusions–. The results showed that when we combine CLOSIB with different texture descriptors, the hit rate increases in all the cases, concluding that CLOSIB can be used to enhance the description of texture in a significant number of situations.

Resumen

Técnicas para la mejora de la descripción de textura utilizando información estadística del gradiente de la imagen

Esta tesis propone tres nuevos descriptores desarrollados con el mismo objetivo: utilizar la información estadística extraída del gradiente de la imagen para mejorar la descripción de la textura presente en la misma.

Primero, propusimos un nuevo método, Adaptive Local Binary Patterns con desviación estándar orientada (ALBPS), que introducía información relacionada con la orientación en el conocido método de Local Binary Patterns (LBP). Esta idea se introdujo previamente en Adaptive Local Binary Patterns (ALBP), al utiliza la media y la desviación orientada de las diferencias absolutas de los píxeles para realizar una clasificación más robusta frente a cambios locales en la estructura de la imagen. Nuestra propuesta difiere de la anterior en el sentido de que incluye la información de la desviación estándar a lo largo de múltiples orientaciones en el método de descripción en vez de en el de clasificación.

Al descubrir que ALBPS mejoraba a LBP y a la mayoría de sus variantes, ALBP incluido, decidimos desarrollar un nuevo método basado en ALBPS, al que denominamos LOSIB (Local Oriented Statistical Information Booster). LOSIB ha sido diseñado para mejorar el rendimiento de otros descriptores de textura que generalmente se centran en la descripción de características locales. LOSIB mejora la descripción de textura incorporando la diferencia de los niveles de gris en diferentes orientaciones de la imagen. Específicamente, se utilizó la media aritmética de las diferencias de los niveles de gris. LOSIB presenta una mayor robustez y fiabilidad en diferentes tipos de imágenes y problemas comparado con ALBPS gracias a la inclusión de dos parámetros adicionales que son el radio del vecindario y el número de vecinos considerados.

Finalmente, propusimos CLOSIB (Complete LOSIB). Esta generalización de LOSIB incorpora tres parámetros en su configuración: el orden estadístico del descriptor, el radio del vecindario y el número de vecinos. CLOSIB se define a través de una exhaustiva expresión matemática. Además, presentamos tres variantes diferen-

tes: una de menor coste computacional (H-CLOSIB), un descriptor multi-escala M-CLOSIB y una combinación de ambos, HM-CLOSIB)

La principal característica en común de los métodos aquí propuestos es que recuperan información estadística global de las texturas de una imagen, al contrario de la información que aportan los descriptores locales como LBP. Además, hemos utilizado nuestros métodos en combinación con otros descriptores del estado del arte a través de una fusión temprana de los mismos, lo que se ha traducido en tasas de acierto más elevadas. Particularmente, la concatenación de CLOSIB con métodos basados en Local Binary Patterns (LBP) produce buenos resultados debido a que las características de ambos descriptores son complementarias.

ALBPS se evaluó en tareas relacionadas con el reconocimiento de materiales y la valoración de la vitalidad de los espermatozoides de verraco. En cuanto a la primera aplicación, ALBPS logró una tasa de acierto del 61,47 % en el conjunto de imágenes KTH Tips2-a, superando al resto de métodos evaluados basados en LBP. En relación con el último campo, la combinación de ALBPS y el descriptor global WCF13, que extrae las características de Haralick de la transformación Wavelet, consiguió un F-Score de 0.886. Este es el mejor resultado logrado hasta ahora en este campo según nuestro conocimiento y experiencia.

La eficiencia de LOSIB se comprobó en tres campos de aplicación diferentes: el reconocimiento de materiales, la clasificación de la integridad del acrosoma de los espermatozoides de verraco y la estimación del desgaste de la herramienta. Para el reconocimiento de materiales, se utilizaron dos conjuntos de datos diferentes: KTH-Tips-2a y Brodatz32 para probar la solidez de LOSIB. LOSIB se evaluó junto con LBP y varias de sus variantes más conocidas. Combinado con CLBP (LBP completo), los resultados mejoraron en un 5,80 % en KTH-Tips 2a y en un 7.09 % para el conjunto de imágenes Brodatz. Todos los descriptores probados lograron un mayor rendimiento en ambos conjuntos de datos y utilizando dos clasificadores diferentes al ser combinados con LOSIB. En cuanto a la valoración de la integridad del acrosoma de espermatozoides, se combinaron tres descriptores: 13 características de Haralick obtenidas de la matriz de co-ocurrencia de nivel gris (GLCM) calculada sobre la Transformada Waveled Discreta basada en Haar; una descripción de textura utilizando LBP y una descripción de forma basada en Fourier. Este método, llamado WFLP, combinado con LOSIB obtuvo un 100 % de precisión y un 0,9975 de FScore, pudiendo determinar con exactitud el estado de los acrosomas sin usar ni tinciones ni microscopios de alto coste. Para la estimación del desgaste de la herramienta, la combinación de LOSIB y ALBP alcanzó un 74,06 % de tasa de acierto al clasificar el desgaste de forma binaria.

CLOSIB se ha evaluado en tres aplicaciones diferentes: el reconocimiento de materiales, por medio del conjunto de datos KTH TIPS2-a, el reconocimiento facial, utilizando el conjunto de datos JAFFE, y el reconocimiento de tejidos, mediante la creación y presentación de un nuevo conjunto de datos para la recuperación de te-

jidos en el interior de habitaciones bajo diferentes formas, condiciones iluminación y oclusiones. Los resultados mostraron que al combinar CLOSIB con diferentes descriptores de textura, la tasa de éxito aumenta en todos los casos, concluyendo que CLOSIB se puede utilizar para mejorar la descripción de la textura en un número significativo de situaciones.

Contents

List of Figures IV

List of Tables VIII

Acknowledgements XIII

1. Introduction **1**
- 1.1. Motivation . . . 1
  - 1.1.1. Classification of boar spermatozoa . . . 1
  - 1.1.2. Tool wear condition monitoring in edge profile milling heads 2
  - 1.1.3. Content Based Image Retrieval (CBIR) for textiles in rooms . . 2
- 1.2. Objectives . . . 3
- 1.3. Main contributions . . . 3
- 1.4. Thesis Organization . . . 4

2. State of the art **7**
- 2.1. Texture description . . . 7
  - 2.1.1. Advantages of morphological texture features . . . 8
  - 2.1.2. Morphological series and size-distribution features . . . 9
  - 2.1.3. Distance-orientation distributions . . . 10
  - 2.1.4. Rotation and illumination invariant distributions . . . 11
  - 2.1.5. Size-shape distributions . . . 12
  - 2.1.6. Distance-orientation-size . . . 13
  - 2.1.7. Most commonly used methods . . . 13
- 2.2. Application fields . . . 14
  - 2.2.1. Spermatozoa classification . . . 14
  - 2.2.2. Tool wear monitoring . . . 15
  - 2.2.3. Material recognition . . . 16
  - 2.2.4. Face recognition . . . 17
  - 2.2.5. CBIR for textiles . . . 18

3. Methodology **21**

3.1. Local descriptors 21
3.1.1. Local Binary Patterns (LBP) 21
3.1.2. Adaptive Local Binary Patterns (ALBP) 24
3.1.3. Local Binary Pattern Variance (LBPV) 25
3.1.4. Completed Local Binary Pattern (CLBP) 25
3.1.5. Histogram of Oriented Gradient (HOG) 26
3.2. Global descriptors 26
3.2.1. Wavelet Concurrent Features (WCF13 and WCF4) 26
3.3. Shape descriptors 29
3.3.1. Fourier descriptor 29

4. Datasets **33**
4.1. Material retrieval datasets 33
4.1.1. KTH TIPS2-a 33
4.1.2. Brodatz32 34
4.2. Spermatozoa datasets 34
4.2.1. Vitality 34
4.2.2. Acrosome integrity 36
4.3. Inserts datasets 40
4.3.1. Cutting edge Dataset 40
4.3.2. Edge wear Dataset 41
4.3.3. Wear Region Dataset 42
4.4. Face recognition dataset 43
4.4.1. Jaffe 43
4.5. Textile retrieval dataset 44
4.5.1. TextilTube 44

5. Adaptive Local Binary Pattern with oriented Standard deviation (ALBPS) **47**
5.1. Method 47
5.2. Material recognition 48
5.2.1. Experimental Setup 48
5.2.2. Results 48
5.3. Vitality sperm classification 50
5.3.1. Experimental Setup 50
5.3.2. Results with kNN 51
5.3.3. Results with SVM 53
5.4. Conclusions 58

6. Local Oriented Statistical Information Booster (LOSIB) **61**
6.1. Methodology 61
6.2. Material recognition 63

6.2.1. Experimental Setup 63
6.2.2. KTH TIPS2-a dataset results 64
6.2.3. Brodatz32 dataset results 67
6.3. Acrosome integrity classification 69
6.3.1. Experimental setup 69
6.3.2. Results 71
6.4. Tool wear estimation 75
6.4.1. Experimental Setup 75
6.4.2. Cutting edge and Edge wear datasets results 75
6.4.3. Wear region dataset results 76
6.5. Conclusions 80

7. Complete Local Oriented Statistical Information Booster (CLOSIB) 83
7.1. Methodology 83
7.1.1. CLOSIB variants 85
7.2. Material Recognition 88
7.2.1. Experimental Setup 88
7.2.2. Results 88
7.3. Face Recognition 92
7.3.1. Experimental Setup 92
7.3.2. Results 93
7.4. Textile recognition 94
7.4.1. Methodology 94
7.4.2. Performance evaluation metrics 97
7.4.3. Experimental setup 98
7.4.4. Distance measure evaluation 98
7.4.5. Results 99
7.5. Conclusions 100

8. Conclusions and outlook 103
8.1. Work summary 103
8.2. General conclusions 104
8.3. Future research lines 104

9. Conclusiones y perspectivas 105
9.1. Resumen del trabajo 105
9.2. Conclusiones generales 106
9.3. Líneas futuras 106

Bibliography 109

Annex A: Research Activities 121

Annex B: Summary of the dissertation in Spanish

List of Figures

3.1. Local Binary Patterns process over one pixel in gray scale level 22
3.2. Example of one pattern and its 90° rotated pattern. 22
3.3. Example of all the possible orientations of a vector and its LBP value. 23
3.4. $s(g_p - g_c)$ (Equation 3.1) for the 36 unique rotation invariant of a neighbour set of $LBP^{ri}_{8,1}$. 24
3.5. Diagram representing the first level decomposition of a DWT. 27
3.6. (a) First level decomposition of a 2D Discrete Wavelet Transform (DWT) diagram. (b) Image decomposition. 28
3.7. L-point digital boundary represented as a complex sequence of points. 30
3.8. Calculating the $freq_{low}$ and $freq_{high}$ scores from $\overline{d}$. 31
3.9. Illustration of the FSD for an acrosome-intact and an acrosome-damaged spermatozoa head. 32

4.1. Examples of some images of the KTH TIPS2-a dataset under different scales and illumination. From top to bottom: Brown bread, cotton, wool and lettuce leaves. 33
4.2. Brodatz dataset examples of each class. 34
4.3. Two pairs of images captured. In the left of each pair the gray scale images and in the right the fluorescent ones. 35
4.4. The first row shows registered gray scale heads and the second row their range filtered outputs. The first four images are examples of alive heads and the last four of dead ones. 35
4.5. Positive phase contrast image and fluorescence illuminance snapshots of the same boar sperm sample. The intact upright spermatozoa cannot be seen under flourescence illumination which allows us to correctly label each cell. 37
4.6. Boar spermatozoa heads with intact acrosome. 39
4.7. Boar spermatozoa heads with damaged acrosome. 39
4.8. Insert dataset example. 40
4.9. Genie M1280 (left side), lateral view (center image), back view (right side). 40
4.10. Example of some images of the Cutting Edge dataset. 42
4.11. Example of some images of the Edge Wear dataset. 43

4.12. Example of the Cutting Edge divided into regions 44
4.13. Examples of the Wear Regions dataset. 44
4.14. Example of different images of the JAFFE dataset. 45
4.15. An image example of each class in TextilTube dataset. 45
4.16. In rows, images of the same textile class in TextilTube dataset. The yellow rectangles overlaid in the images indicate the bounding boxes of the textile regions of the ground truth. 46

5.1. Results achieved on the KTH tips2-a dataset using different classifiers. 49
5.2. Results achieved on the KTH tips2-a dataset using different descriptors and the weighted kNN classifier. 49
5.3. Results achieved on the KTH tips2-a dataset using weighted kNN and different values of k and distance metrics. 50
5.4. Scheme of the experimental setup. 52
5.5. Results achieved on the boar sperm dataset using different descriptors and the simple kNN classifier. 52
5.6. Results achieved on the boar sperm dataset using ALBP, ALBPS and the original LBP. 53
5.7. Results achieved on the boar sperm dataset using kNN with different k values and distances. 54
5.8. Performance of different global texture descriptors (left) and performance of local texture descriptors LBP and ALBP using $R = 1$, $P = 8$ and $R = 2$, $P = 16$ as neighbourhoods (right). 55
5.9. Performance of our proposed texture descriptors and the best of previous local texture descriptors ALBP2,16 57
5.10. Performance of hybrid global+local texture descriptors. 57

6.1. Extraction of the absolute difference of gray-level values for three pixels to compute LOSIB(1,8). 62
6.2. Example of a histogram of all the absolutes differences along the orientation 0° (thus, $p = 0$) and the mean value used to yield the final LOSIB. 63
6.3. Different orientations using 8 neighbours. 63
6.4. Results on the KTH-TIPS 2a dataset using the global descriptors and the combination with the LOSIB(1,8), the LOSIB(2,16) and both of them. 65
6.5. Improvement in terms of hit rate of commonly used global descriptors in the last years when they are combined with LOSIB. . . 65
6.6. Results on the KTH-TIPS 2a dataset using the LBP descriptors and the LOSIB(1,8), LOSIB(2,16) and LOSIB(1,8)+LOSIB(2,16). 66
6.7. Improvement in terms of hit rate of commonly used LBP-based local descriptors in the last years. when they are combined with LOSIB. . 67
6.8. Results using kNN on the KTH-TIPS 2a dataset. In the left, the global descriptors and in the right the local LBP family. 67

6.9. Results using SVM on the Brodatz32 dataset. In the first table, the global descriptors and in the second, the local LBP-based family. . . . 68
6.10. Scheme of our cross validation experimental setup. 70
6.11. Results achieved with each evaluated descriptor. 72
6.12. Results achieved with WFLP16 + LOSIB(2,16) enhancer. 73
6.13. Performance of the evaluated methods with the best parameters. . . 74
6.14. Results using low-medium-high classification on Edge wear and Cutting edge datasets. 76
6.15. Results using low-high classification on Edge wear and Cutting edge datasets. 77
6.16. Results using SVM and classical LBP and CLBP with different neighbourhoods for the binary classification (H-L). The horizontal lines represent the results achieved with LOSIB. 77
6.17. Results using SVM and classical LBP and CLBP for the ternary (H-M-L) classification. The horizontal lines represent the results achieved with LOSIB. 78
6.18. Results using SVM and fusion of LBP and CLBP with LOSIB for the binary (H-L) classification. 79
6.19. Results using SVM and fusion of LBP and CLBP with LOSIB for the ternary (H-M-L) classification. 79

7.1. Absolute differences of the gray values $d_p(x_c, y_c)$ for $P = 8$ orientations in a neighborhood of radii $R = 1$ and $R = 2$. 84
7.2. Schema that shows the neighborhood around a given pixel. 86
7.3. (a) Circumference that represents the neighborhood considered for the computation of CLOSIB with $P = 8$.(b) and (c) Schemas that represent the computation of CLOSIB$_{8,1,\theta}$ for two different images. . . . 86
7.4. Schemas of the computation of CLOSIB$_{8,1,\theta}$ (a) and H-CLOSIB$_{8,1,\theta}$. . 87
7.5. Hit rates when we describe KTH Tips2-a images with different CLOSIBs and LBP-based descriptors. 89
7.6. Hit rates obtained with a given LBP-based descriptor (LBP, ALBP, LVPV and CLBP) and the concatenations of the descriptor with CLOSIB variants. 90
7.7. Hit rates for the LBP-based descriptors (LBP, ALBP, CLBP and LBPV) and their multi-scale versions. 91
7.8. Hit rates obtained with the concatenation of multi-scale LBP-based descriptors and HM-CLOSIB. The horizontal line represents the hit rate of HM-CLOSIB descriptor. 92
7.9. Results using the concatenation of LBP-based descriptors with CLOSIB variants (CLOSIB, H-CLOSIB, M-CLOSIB and HM-CLOSIB) on the original images of JAFFE dataset. 93

7.10. Results using the concatenation of LBP-based descriptors with CLOSIB variants (CLOSIB, H-CLOSIB, M-CLOSIB and HM-CLOSIB) on the cropped images of JAFFE dataset. 94
7.11. Scheme of the TBIR system. 94
7.12. Scheme of the voting procedure to determine the best distance measure. 98
7.13. Results of the voting process in parts per unity for the different distance measures. 99
7.14. Precision at n (p@n) for all texture descriptors using Correlation distance and $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$. 99
7.15. Success at n (s@n) for all texture descriptors using Correlation distance and $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$. 100
7.16. Arithmetic mean of precision and success at n for intervals of n from 1 to 10, from 1 to 20 and from 1 to 40. 101

List of Tables

5.1. Hit rate in % using weighted kNN with several distances on the KTH tips2-a dataset. 50
5.2. Hit rate in % using kNN on the sperm dataset. 54
5.3. Performance of global descriptors 54
5.4. Performance of local texture descriptors LBP and ALBP using $R = 1$, $P = 8$ and $R = 2$, $P = 16$ as neighbourhoods. 55
5.5. Performance of our proposed texture descriptors compared with the best of previous experiments ALBP2,16. 56
5.6. Performance of hybrid global+local texture descriptors. 58

6.1. Performance, in %, of the combinations of LOSIB with the global and local descriptors on the KTH-tips 2a dataset when classifying with SVM. 68
6.2. Performance, in %, of the combinations of LOSIB with the global and local descriptors on the Brodatz32 dataset when classifying with SVM. 68
6.3. Results obtained for the evaluated descriptors showing just the best distances (d) in the GLCM (Gray Level Co-occurrence Matrix). 74
6.4. State of the art results in spermatozoa classification based on the acrosome integrity state. 75

7.1. Each row describes the parameters used to compute different CLOSIBs and H-CLOSIBs in the experiments. Both values, 1 and 2, in the column 'order' indicates that we obtained CLOSIB as a concatenation of the CLOSIBs for each statistical moment, $CLOSIB_{P,R,1\,||\,P,R,2}$. . . 87
7.2. Each row describes the parameters used to compute different M-CLOSIBs and HM-CLOSIBs in the experiments. Several values for a parameter indicate that we obtained CLOSIB as a concatenation of the CLOSIBs for each single value. 88
7.3. Hit rates (in %) obtained with a given LBP-based descriptor (LBP, ALBP, LVPV and CLBP) and the concatenations of the descriptor with CLOSIB variants. 90
7.4. Hit rates obtained in state-of-the-art papers using KTH TIPS2-a. . . . 92

Índice general

Agradecimientos XIII

1. Introducción 1
- 1.1. Motivación . . . 1
  - 1.1.1. Clasificación de espermatozoides de berraco . . . 1
  - 1.1.2. Monitorización del estado de herramientas de corte en cabezas de torneado . . . 2
  - 1.1.3. Recuperacion de imágenes basada en contenido para textiles presentes en habitaciones . . . 2
- 1.2. Objetivos . . . 3
- 1.3. Contribuciones principales . . . 3
- 1.4. Organización de la tesis . . . 4

2. Estado de la técnica 7
- 2.1. Campos de aplicación . . . 14
  - 2.1.1. Clasificación de espermatozoides . . . 14
  - 2.1.2. Monitorización del estado de la herramienta . . . 15
  - 2.1.3. Reconocimiento de materiales . . . 16
  - 2.1.4. Reconocimiento de rostros . . . 17
  - 2.1.5. CBIR para tejidos . . . 18

3. Metodología 21
- 3.1. Descriptores locales . . . 21
- 3.2. Local Binary Patterns (LBP) . . . 21
- 3.3. Adaptive Local Binary Patterns (ALBP) . . . 24
  - 3.3.1. Local Binary Pattern Variance (LBPV) . . . 25
  - 3.3.2. Completed Local Binary Pattern (CLBP) . . . 25
  - 3.3.3. Histogram of Oriented Gradient (HOG) . . . 26
- 3.4. Descriptores globales . . . 26
  - 3.4.1. Wavelet Concurrent Features (WCF13 y WCF4) . . . 26
- 3.5. Descriptores de forma . . . 29
  - 3.5.1. Descriptor de Fourier . . . 29

4. Conjuntos de imágenes 33
4.1. Conjuntos de imágenes de recuperación de materiales 33
4.1.1. KTH TIPS2-a 33
4.1.2. Brodatz32 34
4.2. Conjunto de imágenes de espermatozoides 34
4.2.1. Vitalidad 34
4.2.2. Integridad del acrosoma 36
4.3. Conjuntos de imágenes de plaquitas 40
4.3.1. Conjunto de imágenes del filo de corte 40
4.3.2. Conjunto de imágenes del desgaste del filo de corte 41
4.3.3. Conjunto de imágenes de áreas de desgaste 42
4.4. Conjunto de imágenes de reconocimiento de rostros 43
4.4.1. Jaffe 43
4.5. Conjunto de imágenes de recuperación de textiles 44
4.5.1. TexilTtube 44

5. Adaptive Local Binary Pattern con desviación estándar orientada (ALBPS) 47
5.1. Metodología 47
5.2. Reconocimiento de materiales 48
5.2.1. Configuración del experimento 48
5.2.2. Resultados 48
5.3. Clasificación de espermatozoides según su vitalidad 50
5.3.1. Configuración del experimento 50
5.3.2. Resultados con kNN 51
5.3.3. Resultados con SVM 53
5.4. Conclusiones 58

6. Local Oriented Statistical Information Booster (LOSIB) 61
6.1. Metodología 61
6.2. Reconocimiento de materiales 63
6.2.1. Configuración del experimento 63
6.2.2. Resultados con el conjunto de imágenes KTH TIPS2-a 64
6.2.3. Resultados con el conjunto de imágenes Brodatz32 67
6.3. Clasificación de la integridad del acrosoma 69
6.3.1. Configuración del experimento 69
6.3.2. Resultados 71
6.4. Estimación del desgaste en máquinas herramienta 75
6.4.1. Configuración del experimento 75
6.4.2. Resultados sobre el conjuntos de imágenes de filo de corte y de su desgaste 75
6.4.3. Resultados sobre el conjunto de imágenes de las regiones de desgaste 76
6.5. Conclusiones 80

7. Complete Local Oriented Statistical Information Booster (CLOSIB) 83
7.1. Metodología . . . 83
7.2. Reconocimiento de materiales . . . 88
7.2.1. Configuración del experimento . . . 88
7.2.2. Resultados . . . 88
7.3. Reconocimiento facial . . . 92
7.3.1. Configuración del experimento . . . 92
7.3.2. Resultados . . . 93
7.4. Reconocimiento de textiles . . . 94
7.4.1. Metodología . . . 94
7.4.2. Metricas de evaluación del rendimiento . . . 97
7.4.3. Configuración del experimento . . . 98
7.4.4. Evaluación de las medidas de distancia . . . 98
7.4.5. Resultados . . . 99
7.5. Conclusiones . . . 100

8. Conclusiones y perspectiva 103
8.1. Resumen del trabajo . . . 103
8.2. Conclusiones generales . . . 104
8.3. Perspectiva . . . 104

Lista de referencias 109

Anexo A: Actividades de investigación

Anexo B: Resumen de la tesis en castellano

Acknowledgements

El camino que lleva a una persona hacia una tesis doctoral no es sencillo. Hay que estar muy concienciado para embarcarse en un proceso largo, difícil y con muchos altibajos. Sin embargo, el resultado final es en la mayoría de los casos muy positivo: Buenos momentos con los compañeros, motivación y crecimiento personal, y un sin fin de buenas noticias ayudan a continuar el camino con paso firme y seguro. Para ello, lo más importante es rodearse de buena gente e instituciones competentes.

Lo primero de todo agradecer al gobierno de España y en concreto al Ministerio de Educación, Cultura y Deporte por permitirme disfrutar de una beca FPU destinada a investigadores que desean obtener el título de doctor.

También quiero agradecer a mi tutor Enrique Alegre de la Universidad de León por darme la oportunidad de llevar a cabo estos estudios bajo su supervisión y guía, por su paciencia y sus consejos y por ayudarme insaciablemente a conseguir la beca que me financió estos estudios. Además, quiero agradecer a Laura Fernández Robles que accediera a ser mi cotutora. Una trabajadora excepcional que ha estado día y noche ayudándome a dar forma y sentido a todas las ideas locas que escribía en un papel.

Una de las cosas que con más cariño recordaré de estos años son los compañeros que de una manera u otra me han acompañado: Diego, gracias por mantener activo el fuego del pelotazo durante todos estos años, algún día lo lograremos. Dani, gracias por enseñarme a proteger mi privacidad. Sin tus consejos mi tesis podría estar en manos de cualquiera. Víctor, gracias por cederme tu sitio en el despacho. Era sin duda el mejor de todos. Y Claudia, porque aunque coincidimos poco me enseñaste muchas cosas que desconocía. Eres del parche!

También quiero agradecer a todos los amigos que me han acompañado y soportado en los peores momentos… Andrea, Jessy, Jonny, Aitor, Rubén, Sabri, Rober, Nahúm, Eva, Ángel, Janira, Jenny… y todos aquellos que de un modo u otro habéis colaborado a que estos años hayan sido así. También agradecer a todos mis compañeros de Xeridia y Tecnosylva por aguantarme como un zombie en estos últimos meses de esfuerzo.

Por supuesto a mi familia. Gracias a mis padres por apoyarme en todo momento en todas las decisiones que he tomado a lo largo de mi vida. Gracias por educarme como lo habéis hecho y por aguantar todas mis malas contestaciones en muchos momentos de estrés. Gracias a mi hermano, Carlos, por marcarme el camino y ayudarme siempre que lo he necesitado.

Y dejando para el final a lo mejor, Maite. Cuando tenga la tesis impresa escribiré tu nombre al lado del mio. Sin ti nada de esto habría sido posible. Si no fuera por tu perseverancia y tu inteligencia yo nunca habría sido doctor. Gracias por hacer que estudiar todas y cada una de las asignaturas de la carrera fueran un placer. Gracias por incitarme a realizar la tesis doctoral y gracias por estar ahí día tras día. Siempre serás mi compañera favorita de estudios, de trabajo, de baile, de Singstar, de viajes, de piso y de la vida en general. Nunca cambies y lucha por lo que quieres. Yo seré tu compañero en esas batallas.

Oscar García-Olalla Olivera
León
14th July 2017

Chapter 1

Introduction

1.1. Motivation

Texture analysis is a challenging open problem in computer vision. It refers to a set of algorithms applied to detect and describe spatial variations of the gray level of all the pixels in an image. Nowadays, there are multiple fields that profit from automatic texture description, as it makes processes faster with no need of many qualified staff. In this dissertation we present new efficient methods for texture description of digital images. Later on, we apply and test them mainly in three applications: quality assessment of boar spermatozoa, tool wear monitoring in edge profile milling processes and Content Based Image Retrieval (CBIR) for textile recognition in indoor scenes. We have also evaluated some of our methods for face recognition with the intention to prove their robustness in other applications, but a strong contribution to this field is not provided.

1.1.1. Classification of boar spermatozoa

The task of assessing the quality of semen in Artificial Insemination (AI) processes is indispensable in medicine and veterinary. In this thesis, we deal with the issue of boar sperm quality in the AI industry. Owners of farms are interested in using high quality sperm samples to inseminate the farm gilts. Breeding companies must ensure a minimum level of fertilization potential before selling these products.

In order to determine the quality of the samples, human experts usually take into account the number of dead spermatozoa and also the percentage of heads with a damaged acrosome.

Currently, the assessment of these parameters which determines if a semen sample is suitable for AI is carried out visually, using a microscope and stains to be able to manually count the valuable spermatozoa. This manual process is very expensive because of the high cost of the fluorescence microscope, it is highly computational expensive and may produce errors due to the subjectivity of the expert. Furthermore, the staining process increases the rate of dead spermatozoa which will produce unreliable results.

It would be useful to develop an automatic process to detect damaged acrosomes and dead spermatozoa without using stains neither fluorescence microscopy. This task could be carried out using digital image processing to evaluate gray level images of the samples extracting useful information based on the texture analysis. It would make the process faster, more reliable, and would allow breeding companies to save money, as the requirements would just be a digital camera, a computer, and a conventional microscope.

1.1.2. Tool wear condition monitoring in edge profile milling heads

The development of control techniques for wear detection of cutting tools in machining processes is important in automatic production systems. These tool wear monitoring systems help to reduce costs dramatically, either by decreasing the time spent by an operator to inspect the tool or by helping to prevent breakdowns that an operator can miss. In recent years, search about techniques to automatically estimate tool wear is becoming a key factor due to the introduction of high speed machinery, which causes a great reduction of the tool life. University of León participated in the research project FRESVIDA, funded by the Ministry of Economy and Competitiveness, with the aim of finding a system able to determine the tool life in milling processes using just the information obtained from images of the wear zone.

In this context, the use of artificial vision and in particular, texture analysis helps to develop a precise and efficient system, which can determine the state of the tool in real time. Moreover, such system will not be affected by problems related to indirect methods such as the noise that can influence sensors that measure force, vibration, etc.

1.1.3. Content Based Image Retrieval (CBIR) for textiles in rooms

A textil retrieval system can be used as a tool to fight against child pornography. Retrieving one or several textiles of interest that are present in a scene on huge datasets of proven cases of child pornography allows to find images and videos that might be related to the case of interest. Usually, criminals hide their face or significant objects from pictures in order to avoid their identification, but they do not pay attention to the rest of the useful information that can be extracted from the image. In this context, texture recognition has a great advantage over object recognition which is that textiles are rarely removed or changed from the scene. Therefore, textiles are more reliable for scene retrieval.

1.2. Objectives

The main goal of this dissertation is to evaluate the prevalent texture description algorithms in the scientific community and to develop new texture descriptors which outperform the state-of-the-art in diverse applications. Given the previous idea, we defined the following particular objectives:

1. To improve the success rate of well known texture descriptors for the recognition of materials using challening datasets such as KTH TIPS2-a or Brodatz.
2. To increase the effectiveness in spermatozoa classification tasks, i.e. sperm vitality classification and identification of damaged acrosomes using texture description.
3. To automatically evaluate the wear of cutting inserts in milling processes using only the information of their surface.
4. To present a solution for a new and challenging field: content based image retrieval for textiles.

1.3. Main contributions

The main contributions of this dissertation might be summarized as follows:

1. *A new method called Adaptive Local Binary Patterns with oriented Standard deviation (ALBPS) has been proposed.* This method has been evaluated for material recognition and boar spermatozoa vitality with successful results.
2. *A new global texture descriptor named Local Oriented Statistical Information Booster (LOSIB) has been proposed in order to be used in combination with other existing texture descriptors.* A full evaluation has been performed with some datasets for material recognition and recent descriptors in the state-of-the-art.
3. *A new method for describing the state of the acrosome in boar spermatozoa has been proposed. It is based on an early fusion of LOSIB with Fourier information and Haralick features extracted on the Wavelet responses.* This method outperforms previous works and results obtained with individual descriptors.
4. *A generalization of LOSIB, named CLOSIB (Complete Local Oriented Statistical Information Booster) adds new configuration functionalities with respect to LOSIB.* For example, it allows the inclusion of information regarding the mean or the oriented standard deviation of the gray values of an image and multiple neighborhoods. Three variants of CLOSIB have been proposed: H-CLOSIB, M-CLOSIB and HM-CLOSIB. The main differences among these methods and

CLOSIB are the number of orientations considered (Half CLOSIB) and the arrangement of neighborhoods for different radii (Multiscale CLOSIB).

5. *A new CBIR method for textile retrieval of fabrics like carpets, curtains, etc. in indoor challenging scenes has been presented.* Our method extracts local texture description of maximally stable regions (MSER) from the scene images and retrieves the images that contain similar textures.

1.4. Thesis Organization

In this section the structure of this doctoral thesis is described. The first introductory chapter has been focused on the motivation of the dissertation, the main objectives and the best contributions of this dissertation. Now, the remaining chapters of this thesis are organised as follows.

Chapter 2 contains a review of the state of the art about texture description methods. We focus on the application fields which have been included in our research. For this reason, we have described methods for material recognition published in top research journals and conference proceedings. Then, it studies the literature concerning spermatozoa description and classification using computer vision. Moreover, it comments published methods that deal with the advance of tool wear monitoring of the past years. A brief explanation of the last works on face recognition has also been introduced in this chapter. And finally, it reviews texture description methods for CBIR focusing on textiles, which demonstrates the novelty of our application.

Chapter 3 briefly describes some of the most popular methods for image texture description used in the state-of-the-art in the last decades. We have explained the Local Binary Patterns method and some of their most popular variations: ALBP, CLBP and LBPV. Furthermore, we have introduced global descriptors based on Wavelet transform and shape descriptors that have been used in some experiments of this thesis.

Chapter 4 shows all the datasets used in the experiments of this thesis. We have used two dataset for material recognition: KTH TIPS2-a and Brodatz32. We have also introduced two datasets to evaluate the vitality or acrosome state of the spermatozoa. Furthermore, we have described three different datasets used for tool wear monitoring. Then, Jaffe dataset for face recognition has been depicted. Finally, we have presented a new dataset that we created for textile retrieval using images extracted from videos publicly available in YouTube called TextilTube.

Chapter 5 presents our first contribution to texture description called Adaptive Local Binary Patterns (ALBPS) that introduces a modification of the ALPB method. It includes complementary information about statistical information of the gradient over different orientations. In this chapter, we apply the new method in two applications: material recognition and vitality sperm classification.

Chapter 6 introduces LOSIB (Local Oriented Statistical Information Booster) algorithm. LOSIB is a descriptor enhancer based on the extraction of the mean of the gray level differences along several orientations. This method is used in combination with texture descriptors increasing their performance in several experiments in connection with material recognition, acrosome integrity and tool wear monitoring.

Chapter 7 presents a complete explanation of our generalised version of LOSIB, called CLOSIB (complete LOSIB). It incorporates three different parameters into its configuration, i.e. the statistical order of the descriptor, the radius of the neighbourhood and number of neighbours of the neighbourhood. Furthermore, three different variations of the basic CLOSIB method are introduced and compared in the experiment section: Half CLOSIB, Multi Scale CLOSIB and the Half Multi Scale CLOSIB. This method has been evaluated in different applications, such as material recognition, face recognition and also CBIR for textiles.

Chapter 8 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 summary 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

Texture classification is an open problem in the fields of computer vision and pattern analysis. Nowadays, the number of digital images has increased enormously due to the new smartphones and the cheapest cost of cameras production. Multiple companies are trying to increase their productivity integrating automatic processes managed with cameras. Furthermore, the police and other members of security forces have to deal each day with thousands of images where they have to detect a possible crime, a specific object, or connect the crime scenes with criminal acts of the past using walls, carpets, curtains, etc. to identify them. Texture description has been widely used in object recognition problems ((Liang and Juang, 2015), (Zhang et al., 2017), (Goyal and Singhai, 2016)). Moreover, texture description has demonstrate to be useful in processes connected with quality control in factories, pedestrian detection in crowded streets (Castrillon-Santana et al., 2015; Zheng et al., 2017; Shen et al., 2017), or even helping doctors with medical image interpretations ((Liu et al., 2017),(Kavitha et al., 2015)).

2.1. Texture description

Color, shape and texture properties are fundamental for object description. Nowadays, morphological texture description methods are widely used for grayscale and color images description. In the survey presented in (Aptoula and Lefevre, 2011) a great number of texture descriptor are reviewed, ranging from the basic granulometries and pattern spectra to the most advanced multivariate and multidimensional size, shape, orientation and distance distributions.

In (Drbohlav and Leonardis, 2010), textures are described as spatially adjacent basic structures, the placement and orientation of which are governed by certain generation rules. Along with color and shape, texture constitutes one of the three fundamental properties of objects and for this reason, texture has been an object of study for the image analysis community since its very early days.

As it is pointed in (Aptoula and Lefevre, 2011), a great deal of effort has been spent for the characterization of textures, and this produced the emergence of a rich variety of texture analysis tools ((Manjunath and Ma, 1996) (Tuceryan and Jain, 1998)). All of these tools have found applications in many fields, including but not limited to content-based image annotation and retrieval (CBIR), medicine and natural sciences (e. g. cytology, histology MR imagery, etc), geology, petrography, material sciences (e. g. metallography, porous materials, fractography, glass fibers, etc), geography (e. g. remote sensing, digital elevation models, meteorology, etc), oceanography (e. g. sea floor inspection, plankton identifi- cation, etc), industrial inspection (e. g. quality control, defect detection, etc) and biometrics (e. g. face recognition, fingerprints, etc).

In this thesis, our focus is on texture description, where the main objective is to characterize a given texture, usually with the purpose of assigning it to some texture class.

The two main morphological tools used for texture analysis are granulometry and morphological covariance, and both are based on the common principle of morphological series. The first one studies the amount of detail removed by applying successively basic morphological operators along with SEs of various sizes, whereas the latter is the morphological equivalent of the standard covariance operator. They have been both employed successfully in a large number of applications (Soille, 2003)(de Araujo et al., 2015).

According to the survey carried out by (Aptoula and Lefevre, 2011), we can divide the morphological approaches to texture description in five categories: morphological series and size-distribution features, distance-orientation distributions, rotation and illumination invariant distribution, size-shape distributions and distance-orientation-size.

2.1.1. Advantages of morphological texture features

Morphological analysis is very suitable for textures. It is based on spatial relationships between pixels, and has a large number of tools capable of extracting information of size and shape. This becomes especially valuable when focusing on texture primitives. Furthermore, multi-scale morphological tools can be implemented relatively easily, thus managing the primitive texture size variations, while the connected morphological operators represent an additional powerful set of tools capable of exploiting the pixel connectivities. Morphological methods can capture higher order properties of spatial random processes (Baeg et al., 1999), unlike difference statistics and Fourier transforms, which only describe a texture process up to second-order characteristics. And finally, in the case of higher efficiency requirements, the morphological tools can be directly expressed with Boolean algebra, and therefore can be implemented in dedicated hardware.

2.1.2. Morphological series and size-distribution features

Most morphological texture features are based on the principle of morphological series, which lead to one-dimensional or multidimensional distributions, based on one or more SE properties, such as size, shape, orientation, and so on. Applying a morphological filter such as a structural aperture with a SE b of increasing size λ results in a series of successive filtered images with less and less details. To form the final characteristic vector, higher-order statistical moments can be calculated, some of which are usual, including mean, variance, asymmetry, and kurtosis. These characteristics are called granulometric moments (Dougherty and Astola, 1994). In addition, anti-granulometry (also called anti-size distribution) can be similarly measured using series of morphological closure operations rather than apertures. Consequently, while grain size captures bright details on a dark background, the anti-granulometry focuses on dark details on a bright background, providing an effective description of the granularity of a texture.

Structural openings and closures are the most widely used operators to calculate these distributions. However, the use of more sophisticated alternatives such as their reconstruction-based counterparts is also widespread, in fact there have even been initial reports that the latter have a superior noise robustness (Li et al., 1996). The pattern spectrum ϕ, also called pecstrum (Anastassopoulos and Venetsanopoulos, 1991), is a very interesting and classic morphological overall texture feature. It was introduced by Maragos (1989). It can be seen as the morphological counterpart of the well-known histogram. Instead of measuring the distribution of intensities within an image, it is a question of measuring the distribution of sizes (and to a lesser extent, of shapes). To do this, it collects values of the differential series on all the pixels. The spectrum pattern provides an important amount of information related to texture.

Following the attention that these tools received during the 1990s in the context of the description of the texture, several extensions have appeared. One of them was the iterative morphological decomposition (IMD) (Wang et al., 1993), which basically decomposes a given texture into a set of morphologically "simple" functions. In particular, IMD is based on the associated component image s_λ of a pattern spectrum:

$$f_0 = f$$

$$s_\lambda = f_n if \lambda = n$$

$$s_\lambda = (f_\lambda \circ g_{n-\lambda}) if \lambda = 0, 1, ...n - 1$$

$$f_{\lambda+1} = f_\lambda - s_\lambda$$

$$\Sigma_k^n = \lambda = s_k$$

where f is the original image and g_λ is a sequence of SEs. The final features consist of the mean, variance, gradient and area of each of the component images. IMDs have been further developed and equipped with scale invariance (Lam and Li, 1996, 1997).

A sieve is the representation of an image as a graph. One global sieve which captures the granularity of a texture image is the morphological approach. A morphological sieve performs a decomposition by scale. It is sequential, in contrast to the parallel structure of the granulometry. The sieves are described using the notation of the morphology of the graph (Heijmans et al., 1992), and are defined as operations on the connected systems of pixels. At each stage, the sieve operator ϕ removes the extrema of only that scale. Formally, the result of applying the sieve operator ϕ at scale s to an input image f can be written as shown in 2.1:

$$f_s = \begin{cases} \phi_s(f_{s-1}) si\ s \neq 0 \\ f_0 = \phi_0(f) = f si\ s = 0 \end{cases} \tag{2.1}$$

where the ϕ operator is often in the form of a M or N filter, a subset of ASF filters. Moreover, sieves preserve scale-space causality (Bangham et al., 1996) and are invertible transforms, since the original image can be reconstructed up to an additive constant from a simple summation of the granule images g_s, across all scales s. Particle size and by extension pattern spectra are powerful multiscale tools used in the image and particularly in texture analysis but they suffer from a drawback. They do not retain any spatial information and also the position of components removed by each filter is not included in granulometry or in a spectrum of patterns.

This problem was studied by Wilkinson (2002) who suggested the use of a spatially sensitive measure to quantify the intermediate images of a morphological series. It was proposed to replace, for example, the sum of pixel values with spatially invariant moments, thus leading to pattern moment spectra, which can effectively distinguish more images. Ayala and Domingo (2001) also investigated this issue, and proposed a new descriptor called spatial distribution of sizes that constitutes a generalization of granulometry. More precisely, in their descriptor, the filtered images of the morphological series are replaced by their intersection with filtered translated images; The intersection is calculated linearly with a product rather than in a nonlinear form with a minimum.

2.1.3. Distance-orientation distributions

Granulometry and the pattern spectrum rely on openings and closings, however, it is also possible to involve any morphological operator, such as erosion (ξ) to build a morphological texture feature able to describe the global texture content. The morphological covariance feature K is a fine example in this regard. It is defined as the morphological counterpart of the autocorrelation operator. In this case, the

SE b under consideration consists of a set of two points p_1 and p_2 and is defined by both a size 2λ = $|p_1p_2|$ and an orientation $\vec{v} = p_1\vec{p}_2/|p_1\vec{p}_2|$:

$$K^{\vec{v}}(f) = K^{\vec{v}}_{\lambda}(f)|K^{\vec{v}}_{\lambda}(f) = \Sigma_{p\in E}\Pi^{\xi}_{\lambda,\vec{v}}(f)(p) \tag{2.2}$$

where

$$\xi'_{\lambda,\vec{v}}(f)(p) = f(p-\lambda\vec{v}) \bigwedge f(p+\lambda\vec{v}) \tag{2.3}$$

Another definition of the covariance has been given by Serra (1982) where the autocorrelation function is used, thus resulting in the operator ξ' defined by:

$$\xi'_{\lambda,\vec{v}}(f)(p) = f(p-\lambda\vec{v}) \times f(p+\lambda\vec{v}) \tag{2.4}$$

A further generalization of covariance is the variogram of Hanbury et al. (2005), which is of particular interest for color textures.

2.1.4. Rotation and illumination invariant distributions

Although the use of several dot pair orientations transforms morphological covariance into an invaluable anisotropic tool, in practice it is generally preferred to describe a texture irrespective of its orientation.

In Aptoula and Lefevre (2011), they advance to the next stage with a set of new challenges, of which the fully invariant texture description is an important element (Shotton et al., 2009). For this reason, the last decade has shown an increase of the demand of all types of invariances in the context of texture analysis, with respect to scale, point of view and lighting conditions. Although morphological series are inherently suitable for analysis of multiple scales, and morphological scale spaces have been thoroughly studied, rotation and especially invariant illumination description of morphological texture has not received the attention it deserves. Subsequently, a large number of current morphological texture tools lack these otherwise desirable properties.

The authors observed that the standard covariance operator consists of three basic elements, which are open to all kinds of modifications: the structuring element (SE) used that are by default pairs of points, the morphological operator that is by default erosion and evaluation method of processed image, default: the volume of the image, for example, the sum of the values of the pixels. Having established these three directions of exploration, an investigation of each of them has been carried out in Aptoula and Lefevre (2011). Multiple findings are reported in the last reference, formulated as extensions for morphological covariance, such as circular covariance histograms, circular covariance location histograms, and rotational invariant points. All of them are, however, based on the same theoretical concept. In addition, they suggested a combination of all these extensions, called Extended Morphological Covariance (ECOV), which exhibits rotational invariance and illumination. Extensive experiments shown that they exceed many of the best-known texture descriptors for grayscale data.

In Southam and Harvey (2004, 2005a,b, 2009)), an important and relatively recent work on rotation invariant texture description has been carried out. The authors focused on morphological sieves in order to construct morphological series. In particular, they have proposed the Tex-Mex features. Given a texture, the method employs a number of its channels, not necessarily successive granules, obtained through opening and closing filters as well as from M-filters. Then they compute basic and rotation invariant statistical measures such as mean, mean of the absolute value, standard deviation and skewness from them as features. The use of 1D-sieves was also investigated. The sieves were constructed by means of a 1D recursive median filter applied at various orientations. Illumination invariance, on the other hand, concerns the stability of the computed features under varying illumination conditions, such as intensity, illumination source and angle. In this context, morphological operators have a distinct advantage over their linear counterparts. More recently, multiple research groups have been working with texture descriptors with rotation invariance (Ahmadvand and Daliri, 2016; Dharmagunawardhana et al., 2016).

2.1.5. Size-shape distributions

In the standard size-distribution, it is considered a unique parameter λ for measuring the size evolution, through the SE $b\lambda$. This definition, assuming a single size varying parameter λ, prevents us from performing accurate measurements. Indeed, it is not adequate for elliptical or rectangular texture grains for instance, where the two independent axes should be taken into account. All in all, several attempts have been made to build bivariate morphological series, thus allowing to obtain size-shape measurements. In this way, Lefevre et al. (2007)) considered structuring elements with two different size parameters α and β that vary independently. A similar approach has been proposed by Ghosh and Chanda (1998). In their work, conditional parametric morphological operators are introduced, a 2D set of SE with increasing size was built, both on the horizontal and vertical dimensions. From this set of SE they finally computed the bivariate pattern spectrum for binary images. Bagdanov and Worring introduce the same feature under the term rectangular granulometry (Bagdanov and Worring, 2002), while a slightly different definition has been given by Barnich et al. (2006)) to limit the SE to the largest non-redundant rectangles within the analysed object (in binary images). Moreover, to define multiparametric granulometries, a more general expression of m-parametric SE has been used in Gadre and Patney (1992) .

2.1.6. Distance-orientation-size

Covariance extracts a feature vector containing information on periodicity and directionality, whereas granulometry concentrates rather on the granularity of its input. Consequently both are necessary in the general case for an effective texture description. However, their combination is rather ambiguous, as it can be realized in a variety of ways. The obvious method, is to calculate independently each feature vector and then concatenate both of them. Aptoula and Lefevre (2007), proposed an alternative. It consists of unifying the two operators' functionalities by varying in parallel three SE properties: its size, direction and distance. For practical purposes, the erosion operator of covariance is replaced with an opening. Of course, on the contrary of granulometry it is also necessary to employ SE pairs, so that periodicity information may be extracted.

2.1.7. Most commonly used methods

Haralick features have been widely used in the last 30 years for texture description. Recently, in Chaddad et al. (2011) a system founded on Haralick features to detect colon cancer cells has been developed. Similarly, a local Haralick features extraction method was used Ribaric and Lopar (2012) in a palmprint recognition application obtaining very promising results.

Besides, in the last years, methods based on the Wavelet transform have been developed, showing very high performances on texture retrieval problems. In Carbunaru et al. (2009), a system for textile image retrieval was proposed. Independent component analysis (ICA) were applied to the wavelet transform responses achieving average precision rates of 89%-94%. Rakvongthai et al evaluated the performance of Wavelet transform with very noisy images demonstrating the good efficiency of the transform on these environments (Rakvongthai and Oraintara, 2013). Recently, in Florindo and Bruno (2016), a method based on fractals among coefficients and magnitude of a particular wavelet decomposition was proposed. Furthermore, in Bakhshipour et al. (2017), they developed a method for weed segmentation using wavelet multi resolution images and neural networks with very satisfactory results under leaf occlusion or overlapping.

Lastly, local descriptors have become more and more important in the last few years. Concretely, the Local Binary Patterns (LBP) descriptor proposed by Ojala et al. (1994) have been widely used due to their simplicity and high capability to extract the intrinsic features from the textures. Guo and his group have been developing several modifications of LBP such as LBP variance (LBPV) (Guo, Zhang and Zhang, 2010), complete LBP (CLBP) (Guo, Zhang and Zang, 2010) or adaptive LBP (ALBP) (Guo, Zhang, Zhang and Zhang, 2010).

In the last year, multiple methods based on LBP are still being developed such as FbLBP (Pan et al., 2017) or MQLBP (Patel et al., 2016).

2.2. Application fields

Now we are going to study all the different fields of application where our proposals have been demonstrate improvement: spermatozoa classification, tool wear monitoring, material recognition, face recognition and textile retrieval.

2.2.1. Spermatozoa classification

Herd fertility is critical to the prosperity of any breeding animal company. Artificial insemination is both a fertility treatment for humans and a typical method in breeding of pigs or dairy cattle, dominating the reproductive process of many farms. It brings superior species, mainly for human consumption, requiring that mating is performed using viable sperm.

Computer Assisted Semen Analysis (CASA) systems are more and more present in sperm analysis centres (Pascual-Gaspar et al., 2008). CASA system is used to assess the semen quality of many species such as bulls (Farrell et al., 1998), pigeons (Klimowicz-Bodys et al., 2012) or cats (Siemieniuch and Woclawek-Potocka, 2008), not only humans. It is an automatic or semi-automatic and standardised equipment which allows to assess sperm concentration (Vantman et al., 1988), motility (Hirano et al., 2001) or morphology (Wang et al., 2011) in a semen sample (Jiang et al., 2009). However, the major cause of function loss is damage of acrosome due to the leakage of cellular components and inactivation of crucial proteins (Valcárcel et al., 1997). Up to our knowledge, the only way of automatically measuring the state and vitality of a sperm sample is using fluorescence images in combination with Cell Counters or fluorescence microscopy which is a very expensive device. Otherwise, the stained samples have to be manually assessed which requires the use of veterinary experts and specialised equipments leading to an expensive and non-objective task. This work deals with the automatic and reliable assessment of the state of the boar acrosome using phase contrast images, so just a phase contrast microscope, a high-featured digital camera and a computer are needed.

There are few works related with this topic and in most of them the texture description makes up the basic operation. González et al. (2007) proved that the texture features derived from the Discrete Wavelet Transform (DWT) allow to classify the acrosome integrity with an accuracy of 92.09%. Following this idea, Alegre et al. (2008) used the first order statistics derived from the co-occurrence matrix of the image, both computed from the original image and from the coefficients yielded by the DWT. They affirm that this approach outperforms moments-based descriptors (Hu (1962), Lin and Chou (2003) Ruggeri and Pajaro (2002) and Chee-

Way Chonga (2004)) in terms of the classification accuracy they provide yielding the 94.93% which suggests that the edge and some shape features are not enough to describe the state of the acrosome. Besides texture and moments, the gradient magnitude along the contour of the sperm head has been used to describe the acrosome integrity. Learning Vector Quantification (LVQ) has been applied in 2007 with only three prototypes and a hit rate of 83.5% (Petkov et al., 2007) and more recently in 2012 Alegre et al. (2012) with four prototypes reaching a hit rate of 93.2%. Alegre et al. (2013) improved this approach by using 7 inner contours and computing a local texture descriptor for each point of the seven contours and classifying with Relevance LVQ, obtaining the best result so far with a hit rate of 99%. As we can see, global and local texture description of the head of the spermatozoa and contour description of the shape of the acrosome have been considered in previous works. Here we present a comparison and combination of the three approaches with different methods applied in a bigger dataset.

Not only the description of the cells is important but the segmentation of the heads, if any, and the classification step can be critical. Gonzalez-Castro et al. (2009) proposed a novel and intelligent segmentation method based on a changing threshold and on a Watershed with which 90.96% of 763 spermatozoa images have been correctly segmented. Bijar et al. (Bijar et al., 2012) segmented the acrosome of human spermatozoa through a method based on a Bayesian classifier which utilizes the adaptive mixtures method (AMM) and Markov random field (MRF) model to obtain and upgrade the class conditional probability density function (CCPDF) and the a priori probability of each class. As regard to the classification, unsupervised classification methods proved to have better performance than supervised ones (Alegre, Gonzalez-Castro, Suarez and Castejon-Limas, 2009). However besides this experiments, it is very important to yield an accurate classification. González-Castro et al. (2012) and Alegre, González-Castro, Suárez and Castejón (2009) estimate the true and unknown proportion of damaged cells in a sample with help of the Hellinger distance by quantifying the unknown a priori probabilities of test sets.

The assessment of the vitality of a spermatozoon is a close related topic which has also been studied recently. Most of the works are also based on texture description. Sanchez et al. (2005) used the intensity distribution of the cytoplasm densities of the cells whereas Alegre, Garcia-Ordas, Gonzalez-Castro and Karthikeyan (2011) presents a new textural descriptor called NCSR.

2.2.2. Tool wear monitoring

Development of on-line measurement systems to detect the wear level of metal cutting inserts is an issue of utmost importance for the control of automated production systems. These kind of monitoring systems help to reduce production costs, since supervision of an operator is no longer required. This subject becomes increasingly important in modern facilities where machining is performed at very high cutting speeds, which reduces tool life significantly. Consequently, the cost of tool replacement reduces the productivity in a very competitive industry. The last decade has witnessed a growing trend for extending production benefits associated to the use of cutting tools beyond their normal limits. Despite the many efforts focused on this issue, finding a satisfactory solution for on-line monitoring has not been yet achieved due to the great difficulties implied in tool wear measuring.

Metal machining operations such as turning, milling or drilling are processes involved in the manufacture of most of the goods. Although much research has been focused on acquiring advanced knowledge of the behaviour of these processes in the last decades, the complexity of this work in this field makes that this challenge is still alive nowadays (Bhuiyan and Choudhury, 2014). One aspect can be highlighted in this context: the cost of cutting tools represents a significant portion of the final cost, especially in the context of unmanned production. Teti (2002) and Weckenmann and Nalbantic (2003) reported that cost of cutting tools and their replacement account 3-12% percent of total production costs. So, it is clear that in the case of cutting tools, a considerable amount of money can be saved when increasing the efficiency of tool wear monitoring (Kopac, 1998). In addition, about 20% of non-productive time is due to tool failure (Kurada and Bradley, 1997) on modern machines. On-line monitoring of tool wear may improve the decision making process involved in tool replacement based on objective facts, instead of relying on operator's subjective criteria. In particular, important research has been focused in recent years on the use of different sensors for monitoring the condition of cutting tools. Some of the most recent works are (Painuli et al., 2014; Wang et al., 2014; Pratama et al., 2011; Fernández-Abia et al., 2014; Fayad, 2010; Gao et al., 2010; Li et al., 2010; Qian et al., 2010). Use of wear sensors can reduce tool costs in 40% (Lim, 1995), a clear business opportunity.

It is very important, therefore, to develop precise and efficient methods to predict tool wear level, its evolution and its remaining cutting life. A machine control system should be able to provide optimized strategies for tool replacement and for adjusting tool correctors. In this field, the use of computer vision and texture analysis is of interest to determine the wear of cutting inserts.

2.2.3. Material recognition

Material recognition is an important aspect of visual recognition. Even though it differs from texture recognition since one pattern can be made of different materials, texture features are commonly used for material description.

KTH Tips2-a dataset has become a very popular dataset in the last years for material recognition. It was created by Caputo et al and presented in (Caputo et al., 2005). It is one of the most challenging texture recognition dataset due to the long intraclass distance of every materials. For that reason, several researchers are testing their methods with it, as for example, Chen, Tian, Lee, Zheng, Smith and Laine (2010), who proposed a method called MWLD based on the Weber's Law achieving a 64.7% of hit rate on this dataset.Sharma et al. (2012) presented a method called LHS (Local Higher-Order Statistics), which obtains an accuracy of 73.0%. Hussain and Triggs (2012) presented a method called LQP (Local Quantified Patterns) which obtained an accuracy of 64.2% or the method AMBP proposed by Hafiane et al. (2015) based on LBP which achieved a 70.3% of accuracy. As we can see, there is still room for improvement relative to this specific dataset.

Another widely used dataset is the Brodatz32 dataset (Valkealahti and Oja, 1998), which contains gray scale images of 32 textures under rotation and scale attacks, presenting an open problem. We have tested our proposal with these two dataset since we consider that they are quite representative for texture problems.

2.2.4. Face recognition

Face recognition is another interesting field in which texture description has demonstrated to be very useful. This field has become very popular in the last decades impulsed partially thanks to social networks and the incorporation of cameras in phones. Thousands of pictures that contain faces are shared through internet every day. The possibility of recognizing their identity is very valuable for multiple applications. In Elaiwat et al. (2015), proposed a technique based on the Curvelet transform for 3D face recognition. Marsico et al. (2015) developed a method that uses region description to recognize faces after plastic surgery with very interesting results. In this paper, we also evaluate our method for face recognition. In this case, we use JAFFE dataset (Lyons et al., 1998) which was developed by Lyons et al. in 1998 and is yet extensively used for both facial expression and facial recognition tasks. JAFFE dataset is a very challenging dataset due to the high similarity among the women that form the dataset, all of them Asiatics, and to the different gestures adopted. In Mahersia and Hamrouni (2015), authors developed an algorithm based on steerable filters and a bayesian regularization neural network to classify the seven universal emotional expressions which appear in JAFFE. In Shenoy et al. (2015), authors proposed a new technique for face recognition based on a fusion of Wavelet and Fourier features. The evaluation procedure was carried out using several face databases including JAFFE. In the same line of work, Wan et al. (2015) achieved a 79% of hit rate using a new method called Q-SVD+RWN. Zang et al. (2012) yielded the highest accuracy equals to 86.42% by means of EPP algorithm.

2.2.5. CBIR for textiles

CBIR (Content based image retrieval) is one of the most researched techniques in the last decades. The high number of images and videos has created the necessity of extract useful information for all of this kind of multimedia formats. The main interest of CBIR methods consist of retrieving images from a collection which match a user's preference. That images can be described using features, such as colour, shape, texture or any other kind of information that can be extracted automatically from them. Several researching fields such as medical image (Faria et al., 2015; Srinivas et al., 2015; Bugatti et al., 2014), biological (Feng et al., 2016; Mallik et al., 2010) or biometric security (Iqbal et al., 2012), have been achieved interesting results using content based image retrieval techniques.

Object recognition for CBIR deals with the necessity of find objects in the images. Several researching groups are trying to deal with object recognition using different approaches such as template matching (Shih and Yu, 2015; Tan et al., 2016), color (Li et al., 2015; Zhu et al., 2015), or invariant local features (SIFT, SURF, etc.)(Chang et al., 2012; Chen et al., 2015). However most of these techniques fail when the object is too plain or deformable, when there are color modifications due to illumination changes or when there is no enough keypoints to describe it.

This problems usually appears when we are dealing with textiles instead of object retrieval. Same textiles can appear in the images in multiple conditions and with multiple different shapes. For that reason, classical object recognition systems have bad performance when we try to retrieve these kind of objects. There is some similarity between textiles and material retrieval. In the work carried out by Brilakis et al. (2006), a new method for material recognition was proposed based on intensity, color distribution and responses to filter banks features. The results were promising but just using materials with a high homogeneity, which allows an easy segmentation based on canny edge detection, and with high amounts of pixels per region.

Recently, in Zhu and Brilakis (2010), authors presented a new system for detect concrete in images. To do that, they proposed a system for segmenting the image into regions taking into account their colour. After that, they describe the regions using a color descriptor and trained a machine learning classifier to determine if the region contains concrete or not. This method requires color images and can not deal with very heterogeneous regions due to the way of create the image partitions.

In Son et al. (2014), authors proposed a method based on ensemble classifiers in order to distinguish between concrete, steel and wood. One of the main disadvantages of this work is the necessity of big and uniform areas of the same material in the image due to the segmentation process, which consists of divide the original image in subregions of a fixed size. If the material region is smaller than the grid division, lot of information of the background is processed as a material resulting in a not accurate description.

Besides all of these material papers, there are also some textile retrieval works in the state of the art. In Bashar et al. (2005) proposed a system based on three wavelet-domain based features called symmetry, regularity and directionality. On this paper, the authors demonstrate the better performance of the combination of the three features against just each of them isolated using two datasets formed by 150 and 300 images of curtain patterns. Similarly, in 2009 Carbunaru et al. (2009) proposed a method which applies Independent Component Analysis (ICA) over wavelet-domain images. On that case, the researchers chose a dataset composed of images of 30 different fabrics, obtaining an average recognition rate of 94.86%.

Recently, in Chun and Kim (2013), a new method which uses composite feature vectors of color from spatial domain and texture from wavelet-transformed domain is proposed. In contrast with the other papers described before, Chun et al carried out a retrieval system using a large dataset composed by 1343 textile images.

In Huang and Lin (2014) proposed a system based on the combination of color, texture and shape features in order to retrieve textiles over more than 4000 images downloaded from Globle-Tex Co. The retrieval system was based on a signature process extracted by different k-means clusters achieving a 83% of success with the best configuration.

All of these material retrieval systems works great with construction purposes but will not achieve good performance in our textile retrieval problem. The main two differences are firstly, the size of the regions in the image. In our case, textiles can be formed by a few pixels in the image. And secondly, the heterogeneity of the textiles compared with the construction materials which are typically very similar along all their surface.

Multiple research groups are trying to find the best algorithm for region segmentation. In 2016, Zheng and Sarem proposed a method called NAMES which is based on the idea of packing pixels with a very high performance in terms of time (Zheng and Sarem, 2016). In Yang et al. (2015) proposed a method based on color histogram segmentation using HSV color space. In Matas et al. (2004) proposed a method based on the extraction of maximal stable regions taking into account a binary threshold which varies along all the gray scale spectre.

Chapter 3

Methodology

In this chapter we are going to briefly explain some literature methods used along the dissertation for combination with our proposed methods.

3.1. Local descriptors

3.1.1. Local Binary Patterns (LBP)

Local Binary Patterns (LBP) (Kashyap and Khotanzed, 1986) is a gray-scale texture operator that extracts pixel-wise information of an image. The relationships between a pixel and its neighbourhood pixels are described in a way that all neighbours that have values higher than or equal to the value of the central pixel are multiplied by 1, and all the rest are multiply by 0. After that, the LBP for that pixel is extracted by summing up all those values. Equation 3.1 express this concept mathematically and figure 3.1 shows the extraction of one pattern for one pixel.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \ , \ s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{3.1}$$

where g_c are the value of central pixel, g_p the value of its neighbour p, P the number of neighbours and R the radius of the neighbourhood.

Despite LBP is gray scale invariant, rotating a binary pattern results in a different LBP value (see figure 3.2).

In order to achieve invariance to rotation by assigning a unique identifier to each rotated LBP, Equation 3.2 was defined by Ojala et al. (2002):

$$LBP^{ri}_{P,R} = min\left\{ROR(LBP_{P,R}, i) | i = 0, 1, ..., P-1\right\} \tag{3.2}$$

where $ROR(x, i)$ performs a circular bit-wise right shift on the P-bit number x i times. In this way for example the patterns 01001111 and 00111101 will lead to the same output value. An example can be show in figure 3.3.

As mentioned by Ojala et al. (2002) in the same work , some LBP appear to be fundamental properties of local textures and they are called “uniform” patterns

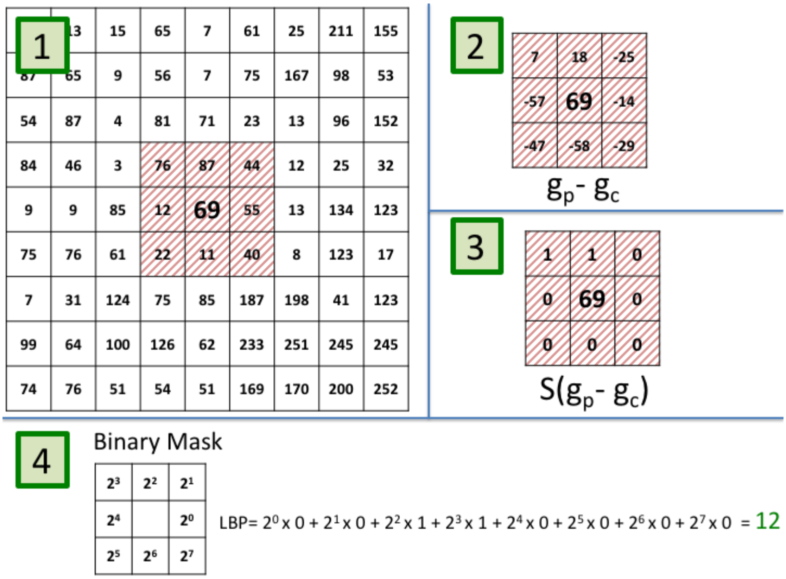


Figure 3.1: Local Binary Patterns process over one pixel in gray scale level using a neighbourhood of radius 1 and 8 different orientations.

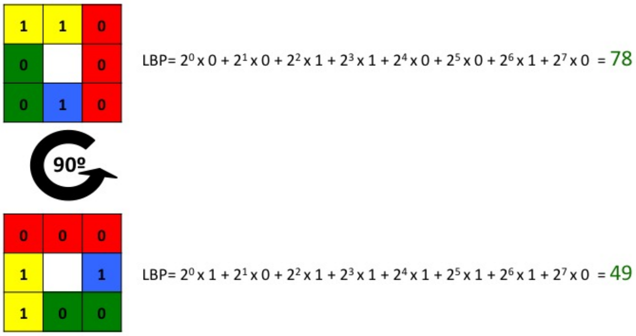


Figure 3.2: Example of one pattern and its 90^{o} rotated pattern.

since they contain very few spatial transitions. Formally, U("pattern") is a uniformity measure which is equal to the number of bitwise transitions from 0 to 1 or from 1 to 0 taking into account also the possible transition between the last and first bits

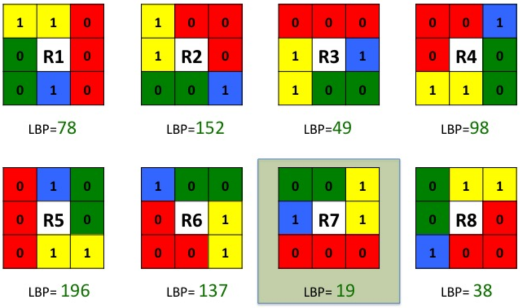


Figure 3.3: Example of all the possible orientations of a vector and its LBP value. In the method proposed by Ojala, the minimum value will be chosen to represent the central pixel.

of the pattern. An LBP is considered uniform if U(“pattern”) is at most 2. Examples of uniform patterns are 00000000 (0 transitions), 00000001 (2 transitions), 00111000 (2 transitions) whereas 01011000 (4 transitions) or 00101011 (6 transitions) are examples of non-uniform patterns. The operator defined in Equation 3.3, apart from being gray scale and rotation invariant, gathers non-uniform patterns in the same group:

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if} \quad U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \tag{3.3}$$

where

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{3.4}$$

There are only $P + 1$ uniform patterns in a neighbour of P pixels. Each of these will have a label from 0 to P, according to Equation 3.3, which corresponds to the number of bits equal to 1 in the pattern. On the other hand, all non-uniform patterns will be labelled as P+1. See figure 3.4. Therefore, in total $LBP_{P,R}^{riu2}$ produces $P + 2$ output values. Finally, a histogram of $P + 2$ bins is built in order to describe the whole image by computing $LBP_{P,R}^{riu2}$ for each pixel of the image, yielding the feature vector of the image.

In this work, we have used $LBP_{P,R}^{riu2}$ but for simplicity it will be called LBP henceforth.

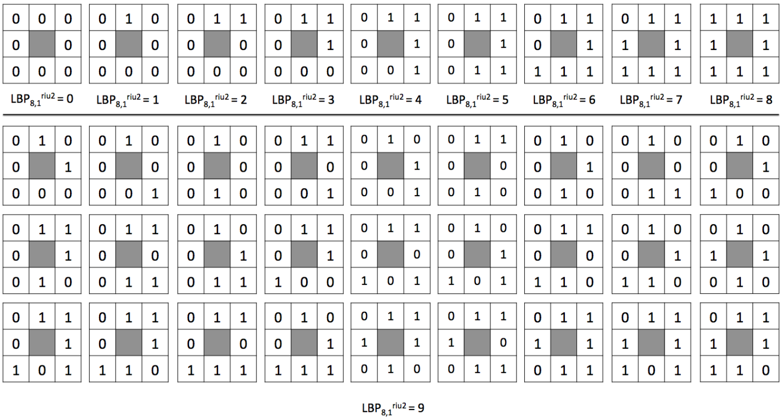


Figure 3.4: $s(g_p - g_c)$ (Equation 3.1) for the 36 unique rotation invariant of a neighbour set of $LBP^{ri}_{8,1}$. First row represents the uniform patterns whereas the rest are non-uniform. Output values for the uniform rotation invariant LBP ($LBP^{riu2}_{8,1}$) are shown under the patterns.

3.1.2. Adaptive Local Binary Patterns (ALBP)

In their study (Guo, Zhang, Zhang and Zhang, 2010), presented an adaptive descriptor based on Local Binary Patterns motivated by the lack of information about the orientation in the Local Binary Patterns approach. In their work, the oriented mean and standard deviation of the local absolute difference are taken into account in order to make the matching more robust against local spatial structure changes. To minimise the variations of the mean and standard deviation of the directional differences, Guo et al. proposed a scheme that minimises the directional difference $|g_c - w_p * g_p|$ along different orientations adding the parameter w. The objective function is defined as follows:

$$w_p = arg_w \min \left\{ \sum_{i=1}^{N} \sum_{j=1}^{M} |g_c(i,j) - w \cdot g_p(i,j)|^2 \right\}, \tag{3.5}$$

where w_p is the weight element used to minimise the directional difference, and N and M are the number of rows and columns in the image, respectively. Each weight w_p is estimated along one orientation $2p\pi/P$ for the whole image.

In their work, Guo et al. used least squares estimation technique to obtain the optimum weight parameter vector (w). After all, the ALBP method is defined as follows:

$$ALBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - w_p \cdot g_c)2^p \ , \ s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \ . \tag{3.6}$$

Similarly to the original LBP, in this dissertation we used $ALBP_{P,R}^{riu2}$ but for simplicity it will be called ALBP.

3.1.3. Local Binary Pattern Variance (LBPV)

LBPV (Guo, Zhang and Zhang, 2010), is a proposal by Guo et al which consists of a combination of LBP and a contrast distribution method. First, the uniform LBP is calculated on the whole image. Then, the variance of the image is used as an adaptive weight to adjust the contribution of the LBP code in the histogram calculation. The LBPV histogram is computed as:

$$LBPV_{P,R}(k) = \sum_{i=1}^{N}\sum_{j=1}^{M} w(LBP_{P,R}(i,j),k), k \in [0,K] \tag{3.7}$$

where k is each bin of the histogram, K the maximum value of LBP and w is defined as:

$$w(LBP_{P,R}(i,j),k) = \begin{cases} VAR_{P,R}(i,j), & LBP_{P,R}(i,j) = k \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

where $VAR_{P,R}$ is the variance of the neighbourhood.

$$VAR_{P,R} = \frac{1}{P}\sum_{p=0}^{P-1}(g_p - u)^2 \tag{3.9}$$

where u is the mean over the different neighbours: $u = 1/P \sum_{p=0}^{P-1} g_p$. On this thesis, we have used $LBPV_{P,R}^{riu2}$ but for simplicity it will be called LBPV.

3.1.4. Completed Local Binary Pattern (CLBP)

In Guo, Zhang and Zang (2010) proposed another method called CLBP to generalize and complete the classical LBP. In that method, a local region is represented by its center pixel and a local difference sign - magnitude transform called LDSMT. LDSMT decomposes the image local structure into two complementary components: the difference signs and the difference magnitudes. In order to code both components, they proposed two operators, CLBP-Sign (CLBP_S) and CLBP-Magnitude (CLBP_M). Since all of them are in binary format, they can be combined to form the final CLBP histogram. CLBP_S is equal to the classical LBP histogram defined in 3.1, and CLBP_M is defined in the Equation 3.10.

$$CLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(m_p, c)2^p \ , \ t(x,c) = \begin{cases} 1 & \text{if } x \geq c \\ 0 & \text{if } x < c \end{cases} \tag{3.10}$$

where c is a threshold determined adaptively.

Finally, CLBP is obtained either concatenating or merging both operators.

On the same way of previous methods, we have used $CLBP_{P,R}^{riu2}$ but for simplicity it will be called CLBP along this dissertation.

3.1.5. Histogram of Oriented Gradient (HOG)

Histograms of Oriented Gradients (Dalal and Triggs, 2005) evaluates local histograms of image gradient orientations over a grid. The main concept of that method is based on the idea that the local appearance of faces, pedestrian or any kind of object can often be well characterized taking into account the local edge directions distribution, even if the corresponding edges locations are not perfectly known. The method is implemented by dividing the image into small uniform regions called cells, often overlapped. Then, for each cell, a histogram of the gradient orientations over the pixels is extracted. The final descriptor is yielded by the concatenation of the different gradients along all the cells of the image. Gradients are computed for every pixel in the cells by convolution with a 1-D gradient filter mask [1 0 -1]. The magnitude of a pixel located in (x, y), (see Equation 3.11), represents the intensity change in the gradient orientation calculated using Equation 3.12.

$$||G(x,y)|| = \sqrt{G_h^2(x,y) + G_v^2(x,y)} \tag{3.11}$$

$$\theta(x,y) = \tan^{-1}(G_v(x,y)/G_h(x,y)) \tag{3.12}$$

where G_h represents the horizontal gradient and G_v the vertical one.

3.2. Global descriptors

3.2.1. Wavelet Concurrent Features (WCF13 and WCF4)

We have used the Discrete Wavelet Transform (DWT) due to its proven efficiency when describing texture patterns. In a simple way, a wavelet transform decomposes an image into the sum of a coarse image plus a detailed image. The DWT of an image

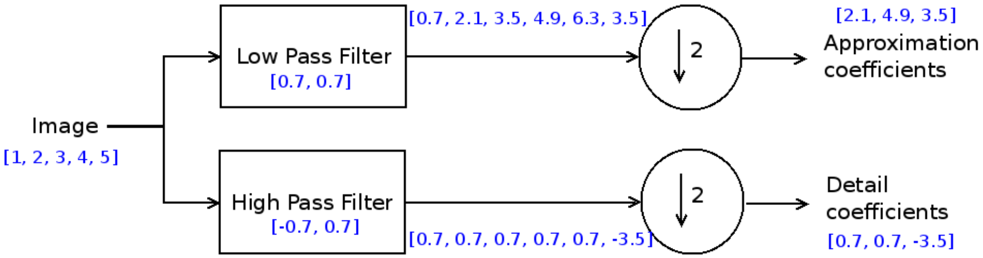


Figure 3.5: Diagram representing the first level decomposition of a DWT. In blue, an example of a DWT applied to a one-dimension vector is shown. The linear convolution between the signal, vector [1, 2, 3, 4, 5], and the low pass filter [0.7, 0.7], results in [0.7*1, 0.7*1+0.7*2, 0.7*2+0.7*3, 0.7*3+0.7*4, 0.7*4+0.7*5, 0.7*5] = [0.7, 2.1, 3.5, 4.9, 6.3, 3.5]. Performing a down-sampling of two, which means that the results in odd positions are deleted, we obtain the approximation coefficients, [2.1, 4.9, 3.5]. The DWT with a high pass filter, [-0.7, 0.7], is computed in the same way.

I is calculated by applying some low and high pass filters g as in Equation 3.13.

$$y[n] = (I * g)[n] = \sum_{-\infty}^{\infty} x[k]g[n-k] \tag{3.13}$$

When the image is passed through a low pass filter, resulting in the lineal convolution between the image and the filter, only the coarse information of the image is preserved. On the other hand, when the image is convolved with a high pass filter, the fine details of the image are kept. This is called level 1 of decomposition. Nevertheless, each course approximation can be filtered, and therefore decomposed, again into a new level of decomposition. After filtering the image, a down-sampling by a factor of 2 is performed, which means that the even elements of the filtered image are preserved while the odd elements are omitted. A diagram together with a simple one dimension example of this approach can be seen in Figure 3.5.

In order to obtain DWT from an image, two dimension DWT has to be considered, leading to a decomposition into four sub-bands in each level of decomposition. These four sub-bands arise from applying horizontal and vertical filters one after the other and they are usually labelled as LL1, LH1, HL1 and HH1 in the first level. The letters refer to the kind of filter applied, L stands for low and H for high, and the number concerns the level of decomposition. For example to compute LH1, first a one dimension low pass filter is applied to each row of the image followed by a down-sampling of 2, leading to an image with half number of rows and the same number of columns. Next, the image is filtered with a one dimension high pass filter and down-sampled with a factor of 2, ending up in an image with both half number of rows and columns as shown in Figure 3.6. The four sub-bands are usually schematised as in Figure 3.6 resulting in an image with the original size. In

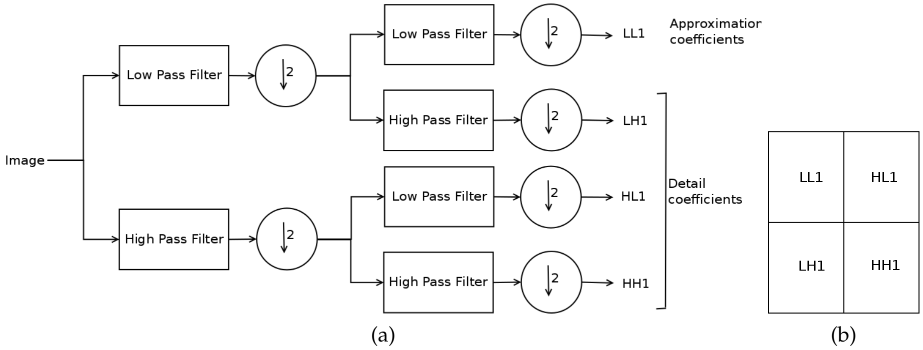


Figure 3.6: (a) First level decomposition of a 2D Discrete Wavelet Transform (DWT) diagram. (b) Image decomposition.

this way, LL1 represents the course information of the image and it is used to obtain the next levels of wavelet coefficients, while LH1, HL1 and HH1 conform the detail images.

We have chosen to use Haar wavelets in the computation of the DWT. The Haar wavelet, introduced by Haar (1910), is the simplest possible wavelet. These set of square-shape functions can be described by its mother wavelet function $\psi\left(t\right)$ as

$$\psi\left(t\right)=\begin{cases}1 & \text{if } 0\leqslant t<1/2\\ -1 & \text{if } 1/2\leqslant t<1\\ 0 & \text{otherwise}\end{cases} \tag{3.14}$$

with scaling function $\phi\left(t\right)$:

$$\phi\left(t\right)=\begin{cases}1 & \text{if } 0\leqslant t<1\\ 0 & \text{otherwise}\end{cases} \tag{3.15}$$

In order to obtain the feature descriptor, some metrics are computed from the GLCM (Gray Level Co-occurrence Matrix) of the original image and the four sub-images of first level of decomposition with the Haar DWT similarly to Arivazhagan and Ganesan (2003).

GLCM 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 the offset $(\Delta x,\Delta y)$ dependent on the direction and the distance at which the matrix is computed, the GLCM is obtained following the Equation 3.16.

$$GLCM_{\Delta x,\Delta y}(i,j) = \sum_{p=1}^{n}\sum_{q=1}^{m} \begin{cases} 1 & \text{if } I(p,q) = i \text{ and } I(p+\Delta x, q+\Delta y) = j \\ 0 & \text{otherwise} \end{cases} \tag{3.16}$$

Some measurements can be computed from the GLCM to characterize an image. Haralick (1979) proposed 14 features. In this thesis, 13 out of those 14 features have been computed, leaving out the Maximal Correlation Coefficient. As mentioned previously, these features are computed on the GLCM of the original image and on the GLCMs of the four sub-bands of the first split of the Haar DWT, leading to a descriptor composed by 65 features which is called WCF13 or composed by 20 features in WCF4 when we just select 4 of the 14 features of Haralick (energy, contrast, correlation and inverse different moment).

3.3. Shape descriptors

3.3.1. Fourier descriptor

Contours are common occurrences in the fields of pattern recognition and image analysis as a way to describe objects. *Fourier shape descriptors* (FSD) provide the means to characterize those contours by capturing either the global features or the finer details of the object's shape. They were first introduced as tools for describing objects in biological samples by Holmquist et al. (1978) and have become popular due to the fact that they are powerful descriptors that are relatively insensitive to noise and easy to normalize (Zhang and Lu, 2003).

Figure 3.7 shows a L-point digital boundary in the xy-plane. By traversing the boundary from an arbitrary starting point $(x(0), y(0))$ and storing coordinate pairs $(x(t), y(t), t = 0, 1, \ldots, L-1)$ the boundary can be represented as a sequence of coordinates $z(t) = [x(t), y(t)]$. Each coordinate pair can also be treated as a complex number so that

$$z(t) = x(t) + iy(t). \tag{3.17}$$

This action has the great benefit that it reduces a 2D problem to a 1D one. At this stage we can choose to eliminate the effect of translation bias by using a shifted coordinate function

$$z(t) = [x(t) - x_c] + i\,[y(t) - y_c] \tag{3.18}$$

where (x_c, y_c) are the centroid coordinates of the object.

For object matching purposes it can be beneficial to sample objects at a fixed number of points rather than using every pixel in the boundary. Using a number of sample points that is lower than the number of pixels in the boundary also

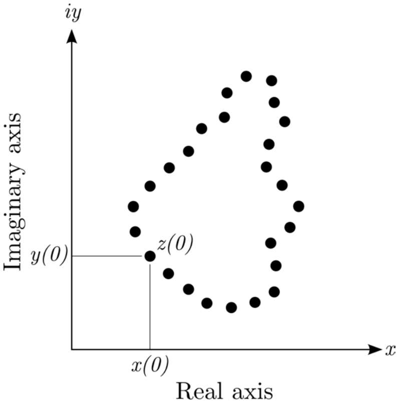


Figure 3.7: L-point digital boundary represented as a complex sequence of points. The arbitrary first point (x_0, y_0) is shown.

smooths the shape, removing noise. There exists different methods for performing the boundary sampling (Zhang and Lu, 2003). For the purposes of this paper the equal arc length sampling method has been chosen. The equal arc length method selects sampling points spaced at equal arclength along the shape boundary. Assuming K sample points the space between two consecutive sample points is given by P/K where P is the shape perimeter.

For a given shape signature, $z(t)$, the discrete Fourier transform (DFT) is written in Equation 3.19

$$u(n) = \frac{1}{L} \sum_{t=0}^{L-1} z(t) e^{-i2\pi nt/L} \tag{3.19}$$

where $n = 0, 1, \ldots, L-1$. The complex coefficients $u(n)$ are called the FSD of the boundary. Fourier shape descriptors are especially good from the point of shape invariants. We are already using a shifted coordinate function (Equation 3.18) that provides us translational invariance. Rotational invariance can be achieved by ignoring the phase information in u, taking only the magnitude values of the FSD. Finally, we can achieve scale normalization by dividing the magnitude values of all descriptors by the value of the second descriptor. An invariant Fourier descriptor vector, $\overline{d}$, can thus be written as in Equation 3.20.

$$\overline{d} = (d_1, \ldots, d_{L-1}) = \left(\frac{|u(1)|}{|u(1)|}, \frac{|u(2)|}{|u(1)|}, \ldots, \frac{|u(L-1)|}{|u(1)|} \right). \tag{3.20}$$

From $\overline{d}$ we can calculate the eccentricity, *ecc*, of the object as in Equation 3.21

$$ecc = \frac{d_1 - d_{L-1}}{d_1 + d_{L-1}} \tag{3.21}$$

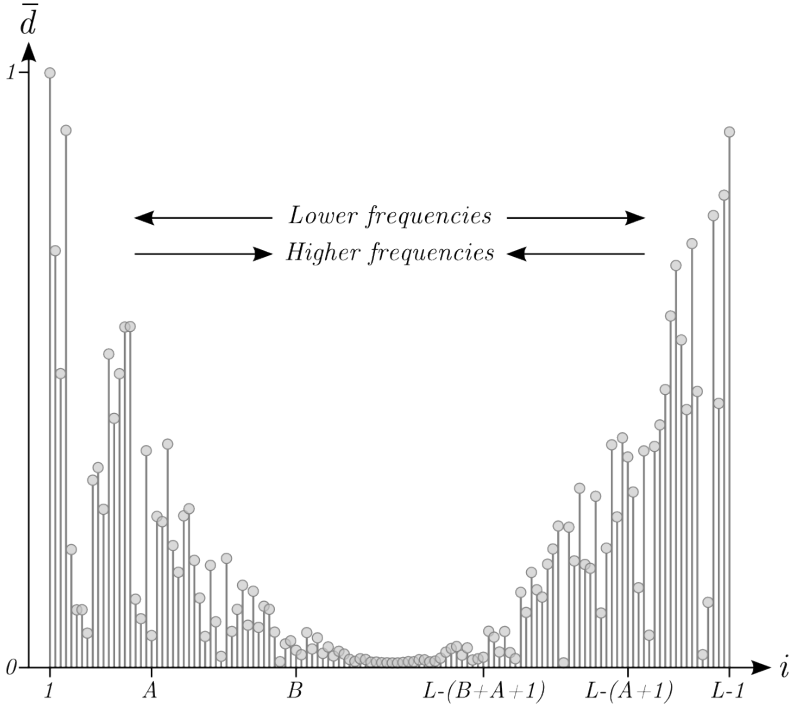


Figure 3.8: Calculating the $freq_{low}$ and $freq_{high}$ scores from $\overline{d}$. Values towards the middle constitute high frequency shape information whereas values towards the edges constitute low frequency shape information.

This constitutes a very rough description of the object shape. To obtain more detailed shape descriptors, it is necessary to locate which intervals of $\overline{d}$ contain interesting information (Rodenacker and Bengtsson, 2003).

We define a lower and an upper limit, $[A, B], B > A$, that will capture the low-, $freq_{low}$, and the high-, $freq_{high}$, frequency shape energy for the object (Figure 3.8). These energy scores are then calculated as following Equation 3.22.

$$\begin{aligned} freq_{low} &= \sum_{i=1}^{A} d_i + \sum_{i=L-(A+1)}^{L-1} d_i \\ freq_{high} &= \sum_{i=A+1}^{B} d_i + \sum_{i=L-(B+A+1)}^{L-(A+2)} d_i. \end{aligned} \tag{3.22}$$

In Figure 3.9, the difference in energy values for different intervals depending on boundary shape is illustrated.

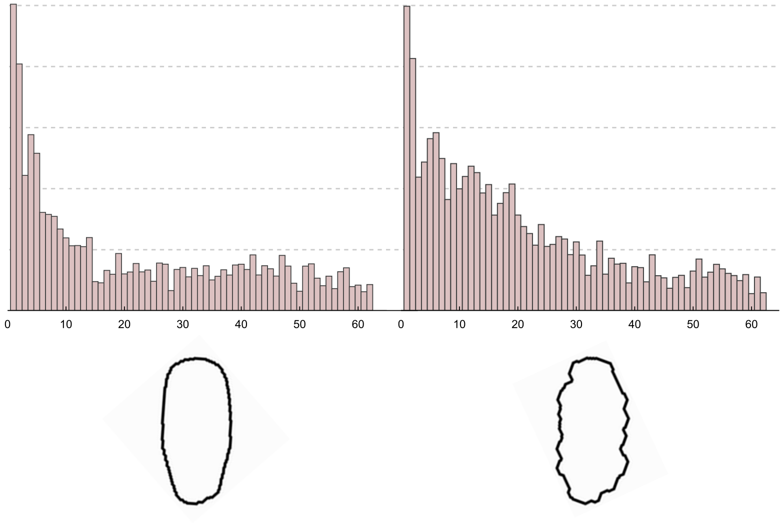


Figure 3.9: Illustration of the FSD for an acrosome-intact and an acrosome-damaged spermatozoa head. The corresponding plots are generated by adding the left half of the spectrum seen in Figure 3.8 with the mirrored values of the right half of the spectrum. The y-axis represents the logarithm of the magnitude of the FSD values. It is clearly discernible, that while the lower frequency values are similar, due to the fact that both shapes are roughly elliptical, and the very high frequencies also similar due to similar amounts of digitization noise on the contour, medium high frequencies show increased values for the unhealthy specimen due to the irregularities of the contour. By binning the energy scores from relevant intervals (see Equation 3.22), this behaviour can be efficiently measured.

Chapter 4

Datasets

4.1. Material retrieval datasets

4.1.1. KTH TIPS2-a

KTH TIPS2-a dataset is composed by 4752 images for material categorization (Caputo et al., 2005). It contains 11 materials (lettuce, brown bread, white bread, aluminium, corduroy, cork, cotton, cracker, linen, wood and wool) with 108 images for four different samples from each material resulting in 432 images per class. All samples were taken at 3 poses, 4 different illumination conditions and 9 scales. It is a very challenging dataset due to all these variations. In Figure 4.1 some examples of textures under different conditions are shown.

Figure 4.1: Examples of some images of the KTH TIPS2-a dataset under different scales and illumination. From top to bottom: Brown bread, cotton, wool and lettuce leaves.

4.1.2. Brodatz32

Brodatz32 (Valkealahti and Oja, 1998) is a subset of 32 images (each image forms a class) of the original Brodatz dataset. It is composed of 2048 sub images of 64×64 pixels (64 images per class) and comprises the following subsets: 16 "original" images, 16 rotated versions of the "original" images, 16 scaled versions of the "original" images and 16 rotated and scaled versions of the "original" images. As a preprocessing step, all the images in the dataset have uniform gray level histogram. In Figure 4.2 we can see examples of each of the 32 "original" images.

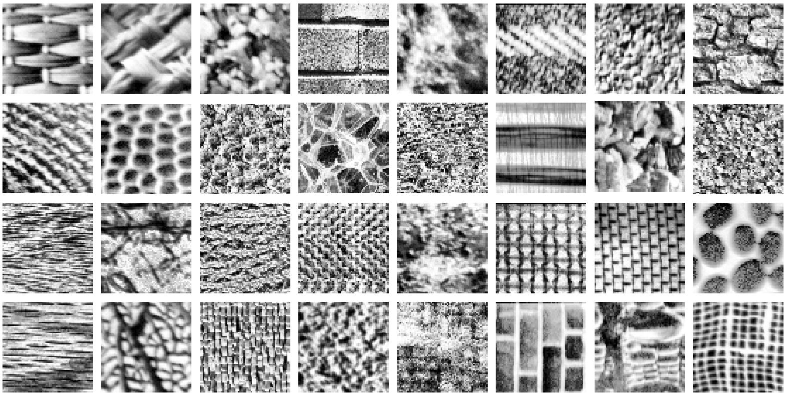


Figure 4.2: Brodatz dataset examples of each class.

4.2. Spermatozoa datasets

4.2.1. Vitality

The lack of publicly available databases of dead and alive boar sperm images drove us to collect an image dataset. This set of images has been captured in CENTROTEC, an Artificial Insemination Center that is a University of Leon spin-off. The sperm was obtained from boars of three different races: Piyorker, Large White and Landrace. 450 pairs of images have been captured using a Nikon Eclipse microscope and a Baster A312f camera of progressive scan. Each of these pairs contains an image in positive phase contrast and a fluorescent image obtained using two different stains: propidium iodide (PI) that dyes dead spermatozoa as red and dichlorofluorescein (DCF) that dyes alive spermatozoa as green. We encourage the reader to find out more about the sample preparation in (Sanchez et al., 2006). We have captured the phase contrast images for developing and testing the texture descriptors evaluated on the proposed method. The fluorescent images were used to obtain the ground truth in order to label all the heads in the dataset. Examples of this captures can be shown in Figure 4.3.

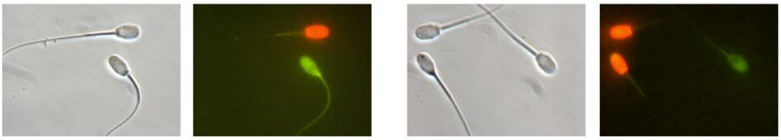


Figure 4.3: Two pairs of images captured. In the left of each pair the gray scale images and in the right the fluorescent ones.

After labelling all the images, each head has been automatically registered in order to assure scale and rotation invariance. First of all, the heads have been rotated to its vertical position. This is performed by relating an sperm head with an ellipse and correcting the orientation of the major axis to achieve verticality. Then, the image has been right and left cropped leaving head's pixels untouched. Afterwards, the tail coordinates have been detected. Evaluating if the tail is placed in the bottom half or in the top half of the image will let us know if the spermatozoon has its head up or down respectively. In the second case, the image has been flipped, leading to equal orientations. Then, the image has been up and down cropped leaving head's pixels intact.

Finally, a 3×3 texture range filter has been applied over the whole dataset in order to reduce the non-informative areas and therefore facilitate the subsequent dataset description and classification. Figure 4.4 shows gray scale dead and alive heads and their filtered outputs.

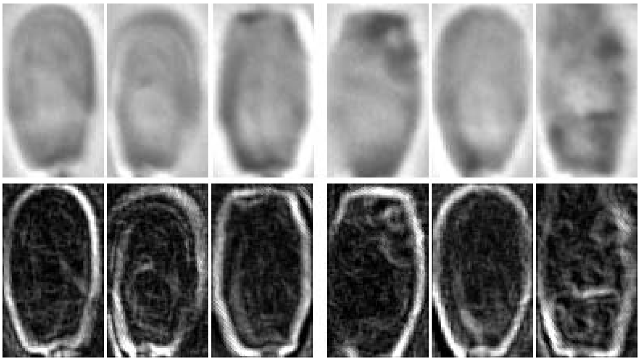


Figure 4.4: The first row shows registered gray scale heads and the second row their range filtered outputs. The first four images are examples of alive heads and the last four of dead ones.

4.2.2. Acrosome integrity

In the same way as we did for the Vitality dataset, we have collected and labelled a boar sperm dataset available at our website[1] for acrosome integrity assessment. The images have been captured at CENTROTEC research centre working together with a group of medical veterinarians. Furthermore, in order to make the dataset valid for a wider variety of boars, sperm samples have been obtained from three races which are Piyorker, Large White and Landrace.

First of all, we have taken an aliquot with 500 μl of diluted sperm. Afterwards, 5 μl of formaldehyde ($<$ 0.3 %), 25 μl of FITC-PNA lectin, and the aliquot were mixed. The formaldehyde paralyses the movement of the spermatozoa in order to make possible a better capture of the images.

The images were acquired using a Basler Scout scA780-54fc camera, with 780 $\times$ 580 pixels of spatial resolution, controlled by a computer in order to adjust its parameters and to store the images. Besides the camera, an epifluorescent microscope Nikon E-600 configured with a $100x$ magnification allowed to capture and to observe a sample under a positive phase contrast and under a fluorescent illumination.

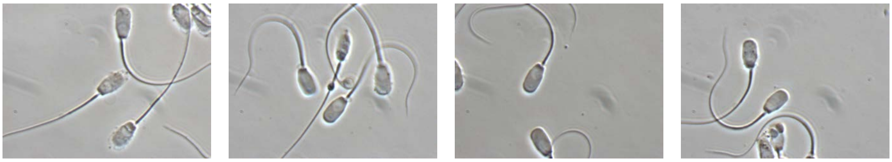
Using this microscope, it was possible to create the ground truth dataset where each head in grayscale is labelled as acrosome-intact or acrosome-damaged according to the color information obtained from its fluorescent paired image. This was possible because when the acquisition had place, two images, one in positive phase contrast and the other one with fluorescent illumination where captured for the same sample. Therefore, for each sample two snapshots were taken and stored, one in positive phase contrast with a conventional light diaphragm and the other one in real color with fluorescence illumination. The positive phase contrast image is a gray level image that was the one used in the image processing, namely the segmentation and description of the spermatozoa heads. On the contrary, the fluorescent illumination snapshot was only used to manually label each sperm head as acrosome-intact or acrosome-damaged in order to create the reliable ground truth. Overlapped heads cannot be analysed, so they are discarded from the set of images.

First row of Figure 4.5 presents several examples of the positive phase contrast images whilst in second row fluorescent images are shown. The samples were stained with hypogaea (peanut) agglutinin (PNA) that is a labelling method where the labelling is restricted to the acrosome and is not influenced by the fixation procedure. As it is known Fazeli et al. (1997), the FITC-PNA binding site is mainly limited to the outer acrosomal membrane of boar sperm, therefore it is an accurate test for studying boar sperm acrosome reaction.

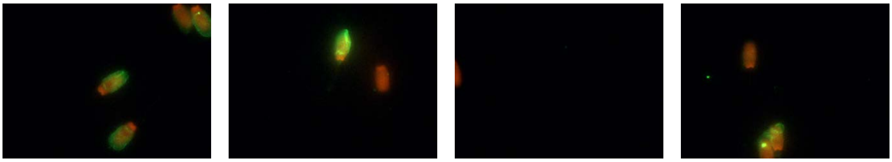
After this labelling, the dataset used was meaningful for classification with supervised learning techniques. More details about the sample preparation can be

---

[1] http://pitia.unileon.es/VARP/datasets



Positive phase contrast



Fluorescence illuminance

Figure 4.5: Positive phase contrast image and fluorescence illuminance snapshots of the same boar sperm sample. The intact upright spermatozoa cannot be seen under flourescence illumination which allows us to correctly label each cell.

found in (Sanchez et al., 2006).

Each of the spermatozoa has been cropped to obtain only one spermatozoon per image. To do that, we used the ideas presented in Gonzalez-Castro et al. (2009) work, where images of alive and dead spermatozoa in positive phase contrast are segmented. We have followed the same steps but modifying the validation criteria in such a way it works when the purpose is to segment acrosome-intact and acrosome damaged spermatozoa heads. Gonzalez-Castro et al. (2009) method combines a first segmentation step, where some morphological operations and Otsu's thresholding is carried out, with a second segmentation using a Watershed transform, in cascade. After the first step, the images that do not fulfil the validation criteria explained below are segmented again and, if they do not accomplish them again, they are rejected.

The criteria for rejecting a wrongly segmented image are based on specific values that have been obtained experimentally and that are the following ones:

1. The area, measured in pixels, of the obtained head cannot be smaller than the 75% of the average area of the whole dataset. This average area was computed using a ground truth where the heads were segmented manually. As long as the camera resolution, and the magnification of the image (microscope and optics used) remain constant, this value is constant.

2. The ratio between the major and minor axes of the ellipse that has the same normalized second central moment as the considered head, must belong to the interval [1.5, 2.4]. These values were also obtained empirically using the same

ground truth dataset.

The first segmentation was carried out preprocessing the images, thresholding them and applying several morphological operations. In the preprocessing, images are converted to grayscale and image contrast is increased with a histogram slicing. Later, a binary image is obtained with a threshold calculated using Otsu's method (Otsu, 1979), the negative image is computed, the white regions are dilated using a disk-shaped structuring element and all the objects smaller than the biggest one are removed. Finally, the spermatozoa tail is separated using an opening and later removed in such a way that only the white region corresponding to the spermatozoon's head remains. This region will be used to mask the grayscale level image if the validation criteria are accomplished (regions properties like eccentricity or area).

When an image does not pass the segmentation test, the previously segmented region of interest, that did not accomplish the validation criteria, is used as the foreground marker for the Watershed transform. The markers are necessary to avoid the typical over-segmentation due to the fact that the gradient of the image is very sensitive to noise. In the non frequent cases when the previous segmentation do not yield any region, a box placed on the center of the image, with an area of the 25% of the image size, is used as the foreground marker. The background markers are obtained by computing the watershed transform of the distance image of the foreground marker to each one of the corners of the image. After performing the watershed segmentation, the regional minima of the gradient image is computed and the area of the foreground marker obtained with the watershed transformation is smoothed using an opening with a disk-shaped structuring element.

Finally, the validation criteria is checked out again, If the segmented image passes the criteria the grayscale image is masked with the binary image and, otherwise, the image is discarded and labelled as wrongly segmented.

Using this method, approximately 10% of the original images are non well segmented and rejected before the description. The good thing about this method, which also was the main reason for using it, is that non even one image wrongly segmented passed to the description step, as was checked by a visual validation of the whole segmented set. In this way, we are rejecting some rightly segmented images but with this approach we do guarantee that the posterior description is accomplished only on rightly segmented images. On the other hand, to discard some spermatozoa is not important in a CASA (Computer-Aided Sperm Analysis) system because usually, depending on the conditions of the sample (dilution, centrifugation, etc.), it is frequent to find overlapping spermatozoa, particles and other circumstances that recommends that the number of the spermatozoa that are present in a sample should be much higher than the number of spermatozoa needed to assess this sample.

Finally, the dataset is composed by a total of 1851 images with 905 intact and 946 damaged acrosomes. This dataset is the biggest one up to date, compared with most of the papers published on this topic. In (Alegre et al., 2013) only 360 acrosomes were used (210 intact and 150 damaged) whereas Alegre et al. (2012) used a total of 800 sperm heads, 400 intact and 400 damaged. On that account, more reliable results are expected with this bigger dataset. Examples of intact and damaged spermatozoa heads can be found in figures 4.6 and 4.7.

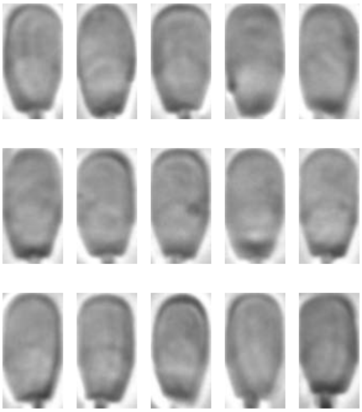


Figure 4.6: Boar spermatozoa heads with intact acrosome.

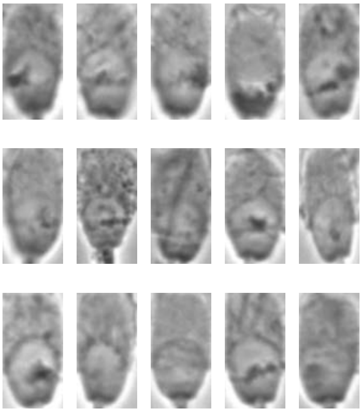


Figure 4.7: Boar spermatozoa heads with damaged acrosome.

4.3. Inserts datasets

4.3.1. Cutting edge Dataset

Characterizing the inserts based on the wear is one of the interest field for texture classification. Because of the lack of a datsaset that fits our requirements, we have acquired an Insert edges dataset processing the images of 53 tools as the ones shown in Fig. 4.8.

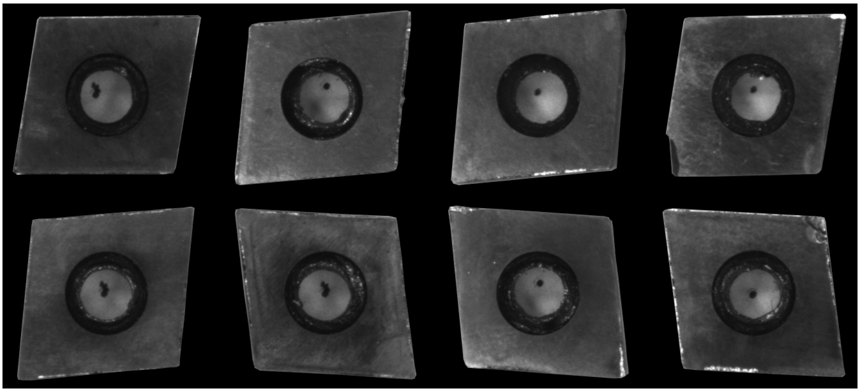


Figure 4.8: Insert dataset example. Inserts on the left present high wear while inserts on the right present low wear taking into account the left cutting edge.

The images were captured using a monochrome camera, model Genie M1280 1/3″ with a $25mm$ optic AZURE. The focus and aperture are manual and the sensor has a resolution of 1280×960 pixels. The monochrome camera is shown in Fig. 4.9.

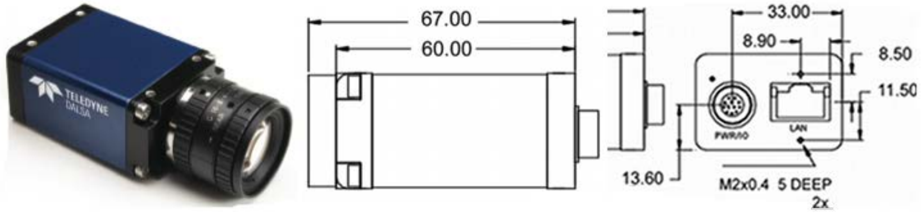


Figure 4.9: Genie M1280 (left side), lateral view (center image), back view (right side).

Central region elimination

The gray-scale images of the inserts with masked background are subjected to a pre-processing step which results in four new images, one for each cutting edge in horizontal position. The first step consists of removing the central portion of the insert, a circular masking region. Firstly, the center and radius of the circle are determined. To determine the circle center, the image is binarized with a threshold of 0.01, resulting in an image with insert pixels set to 1 and the rest to 0. Then, the centroid of the insert area is calculated and assigned as the circle center. Using this binarized image, the length of the major axis is also obtained. The circle radius is defined as 1/5 of the major axis length. After this process, we obtain a gray-scale image of the insert with both its background and central region masked out. This image will be used to extract the four cutting edges.

Edge extraction

This section describes the edge extraction just only for the west edge. Rest of edges (north, east and south) are extracted applying the same procedure on 90, 180 and 270 rotated version of the original image, respectively. A vertical Sobel filter is applied to detect the insert contours using the gray-scale image that results from the previous step. These contours are then dilated and opened, and vertically projected on the horizontal axis. The first element non-equal to zero in this projection indicates the x coordinate of the image where the cutting edge starts. Then, from this starting x coordinate the image is cropped to a width of 100 pixels and the same height of the original image, so that it contains the cutting edge. A parametric margin is added to increase the cropped area because experimental tests showed that some inserts lost some edge pixels due to a tight crop.

Once the four cutting edges have been extracted and their crops have been rotated to horizontal position (edge facing to south), the edges need to be aligned to horizontal position. Because inserts are not square-shaped but rhomboid-shaped instead, their cutting edges are not aligned within the same vertical and horizontal axis. Therefore, a rotation is required to align edge image with regard to the horizontal axis. To carry out this operation, a horizontal Sobel filter is applied followed by a dilation. The resulting image is filtered in order to remove smaller objects which results in a binary image containing just the cutting edge. This edge is inscribed in a ellipse whose major axis is obtained. The orientation of this major axis is the same as the orientation of the cutting edge with respect to the horizontal axis. Later, the original cutting edge image is rotated to compensate this orientation, obtaining cutting edges aligned to horizontal. 212 such images conform the Cutting Edges dataset. An example of four images of this dataset is shown in Figure 4.10.

4.3.2. Edge wear Dataset

In order to describe only the tool wear region of interest instead of all the insert area, a manual segmentation over the worn area was performed using the Cutting

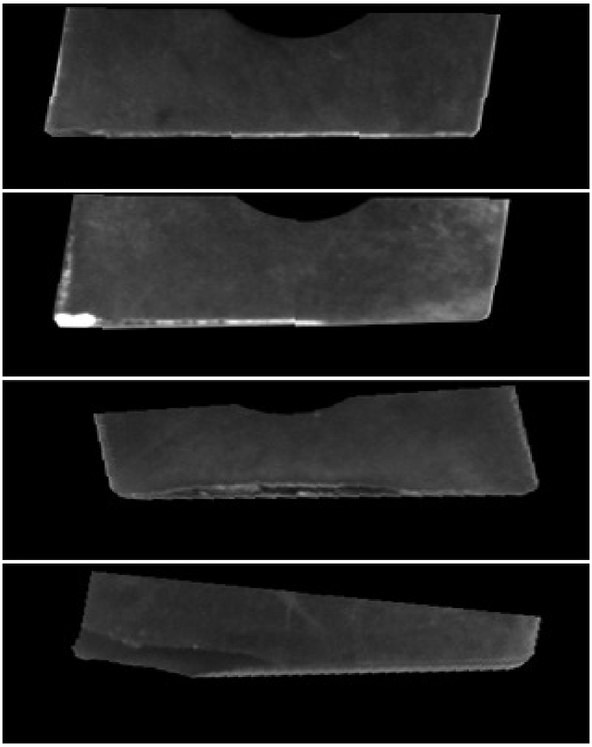


Figure 4.10: Example of some images of the Cutting Edge dataset.

edge dataset. These images were considered as a new dataset called Edge Wear dataset. An example is illustrated in Figure 4.11 for the same images shown in Figure 4.10.

4.3.3. Wear Region Dataset

The grey scale images of the inserts with masked background were subjected to a pre-processing step, which resulted in four images, one for each cutting edge. However, several regions with different level of wear could be determined in the same cutting edge image. An example of extraction of wear is shown in Figure 4.12. As we can see, an image can contain regions with different levels of wear. For this reason we create a new dataset. Now, we aim at classifying homogeneous wear areas independently of the cutting edge in which they are found. The Cutting Edge dataset was formed by 212 images. Extracting the wear regions of each image yields a final dataset of 577 wear tool patches.

Some images of the new dataset are shown in Figure 4.13. In the first row vertical regions are shown. The second and third rows contain horizontal regions.

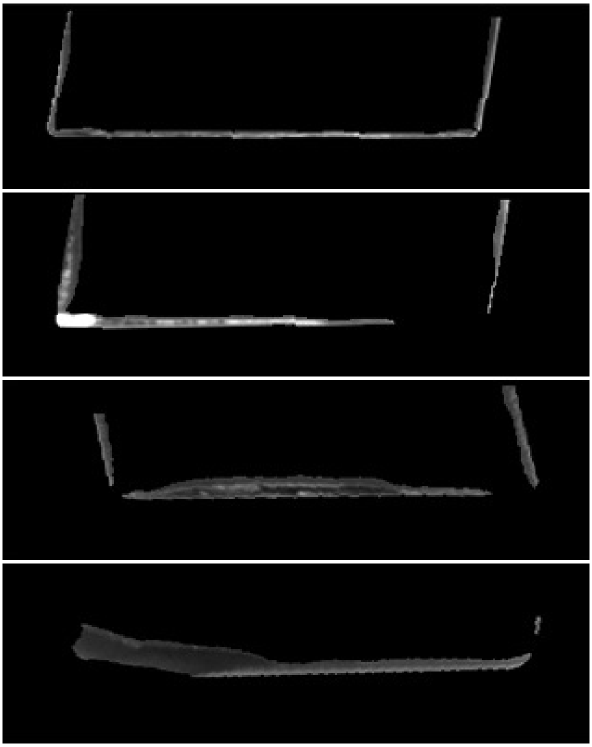


Figure 4.11: Example of some images of the Edge Wear dataset.

4.4. Face recognition dataset

4.4.1. Jaffe

JAFFE dataset is composed of expression images of 10 Japanese female subjects. It was presented in Lyons et al. (1998). All the images were normalized to 256×256 pixels and their were taken under similar light conditions based on tungsten lights. The images were taken from a frontal pose. A preprocessing step was carried out before describing the image. We cropped the image to just take into account the area of the face, avoiding the background information. The dataset contains 213 images with represent the spontaneous expression of anger, disgust, fear, happiness, sadness, surprise and also neutral. Example of different expressions of two subjects is shown in Figure 4.14.

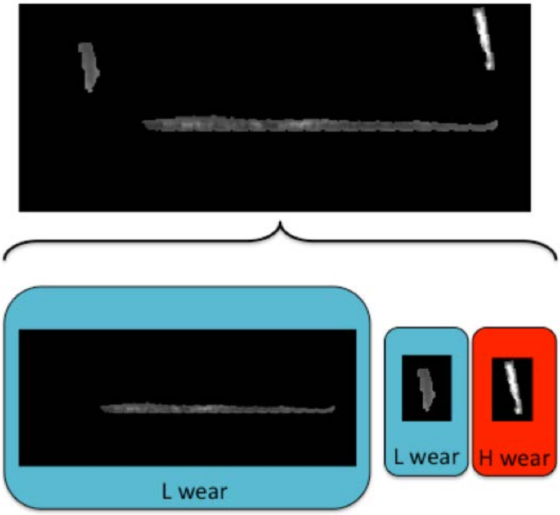


Figure 4.12: Example of the Cutting Edge divided into regions

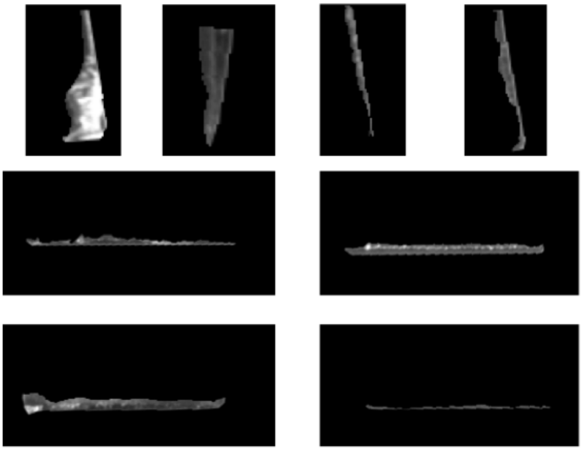


Figure 4.13: Examples of the Wear Regions dataset.

4.5. Textile retrieval dataset

4.5.1. TextilTube

Textile retrieval in real environments is a poorly investigated research field. Up to our knowledge, there is no publicly available dataset that focuses on the recognition of rigid and non-rigid textiles presented in different sizes, shapes and capture

Figure 4.14: Example of different images of the JAFFE dataset. In each row one subject and in each column a different expression example.

conditions. For this reason, we created a new dataset for the retrieval of textiles in bedrooms, which is available in [2].

The dataset is composed of 690 images obtained from 15 videos of YouTube. The videos were recorded in bedrooms with plenty of textiles and with different camera poses, illumination conditions, occlusions, etc., which makes the textile retrieval task very challenging. The dataset contains 67 classes of textils such as curtains, carpets, sofas, shirts, dresses, etc. Figure 4.15 shows a mosaic encompassing one image of each class. The number of elements of each class varies from 4 to more than 100. There is a total of 1787 regions. Therefore, the dataset is highly skewed, simulating a real scenario.

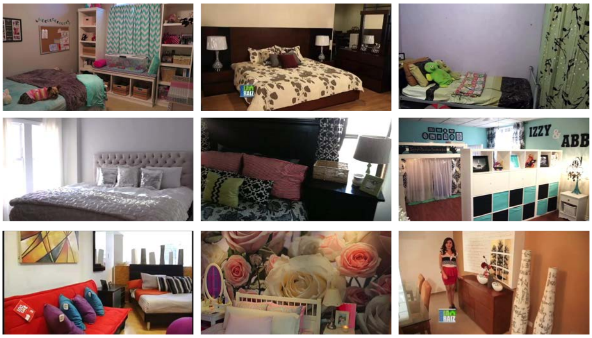


Figure 4.15: An image example of each class in TextilTube dataset.

We labelled the dataset in order to provide a ground truth that allows the user to automatically evaluate the performance of a method on the dataset. The ground truth includes the bounding box coordinates and the class labels of each textile re-

---

[2] pitia.unileon.es

gion in the images of the dataset. We provide the ground truth in the form of an XML file. We show the diversity in terms of type, size, pose, etc. of some textile classes of the dataset in Figure 4.16.

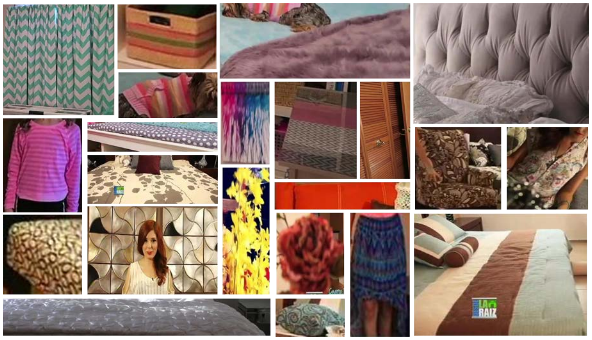


Figure 4.16: In rows, images of the same textile class in TextilTube dataset. The yellow rectangles overlaid in the images indicate the bounding boxes of the textile regions of the ground truth.

Chapter 5

Adaptive Local Binary Pattern with oriented Standard deviation (ALBPS)

5.1. Method

In Guo, Zhang, Zhang and Zhang (2010) proposed the use of the oriented mean and standard deviation in the matching algorithm to improve the classification performance. However, their method, called ALBP, does not take into account these statistical values to compute the image descriptor. Instead they were only applied to minimize the directional difference along the different orientations using the weight parameter w_p. Our proposal includes the standard deviation information not in the matching method but in the descriptor algorithm and it is called ALBPS on that account. Furthermore, whereas a 1 by 1 matching technique was proposed by Guo et al., our scheme allows supervised classifiers training. This is a huge advantage in most cases because, in this way, it is possible to use a fast and powerful classifier that will perform very well when the training set is big enough.

The standard deviation vector σ is obtained using the Equation 5.1.

$$\sigma_p = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}(g_c(i,j) - g_p(i,j) - \mu_p)^2/(M \cdot N)} \tag{5.1}$$

where N and M are the numbers of rows and columns respectively, $g_c(i,j)$ is the center pixel at position (i,j), $g_c(i,j)$ is neighbourhood of $g_c(i,j)$ lying along orientation $2p\pi/P$ with radius R and μ_p the oriented mean obtained using:

$$\mu_p = \sum_{i=1}^{N}\sum_{j=1}^{M}|g_c(i,j) - g_p(i,j)|/(M \cdot N) \tag{5.2}$$

ALBPS descriptor is obtained by concatenating the $P+2$ bins histogram values of the uniform LBP approach together with the P-dimensional standard deviation vector, yielding a descriptor of $2P+2$ features with P the size of the neighbourhood.

5.2. Material recognition

5.2.1. Experimental Setup

In this section, we are going to evaluate the performance of ALBPS over the KTH Tips2-a dataset, which is focused in material recognition as we have explained in chapter 4. The experimental setup used in this experiment is the standard protocol developed by Caputo et al and used in many works (Caputo et al., 2005; Chen, Shan, He, Zhao, Pietikainen, Chen and Gao, 2010). Basically, it takes one of the samples of each material for test and the rest for training, which conforms a more challenging setup than using random images for test and training. In this experiment, we carried out a four cross validation classification to increase its robustness, one classification for each sample in the test class using three k Nearest Neighbours (k-NN) approaches (standard kNN, mean distance kNN and Weighted kNN). As it is known, kNN classifier does not need a training step so the test and training subsets here are used in a different way: all the images of the test subset are compared against the images belonging to the training set. The final hit rate is obtained as the mean of the hit rates in each iteration.

5.2.2. Results

Nearest Neighbours variants evaluation

Three different classifiers based on the kNN method have been implemented in order to evaluate different LBPs descriptors such as LBP, ALBP, LBPV, CLBP and the proposal ALBPS. Four distances have been selected due to their good general behaviour: euclidean, intersect distance, Chi square and Cityblock. In Figure 5.1 we can observe the mean performance along different values of k ($k = 1, 3, 5, 7, 9, 11$) of all the classifiers for each pair of LBP descriptor and distance. As we can see, the best results are obtained using the Weighted kNN for all the test except for the LBP-Chi square in which standard kNN performs better.

LBP Variants Performance with Weighted NN

In Figure 5.2, we can see the mean performance along different k values of the weighted kNN classifier using different distance metrics. It can be appreciated that the proposed ALBPS outperforms all the others methods for all the distances except for the Euclidean metric in which ALBPS and ALBP get the same result. The best performance was obtained using Chi Square and ALBPS with a 60.23% of hit rate.

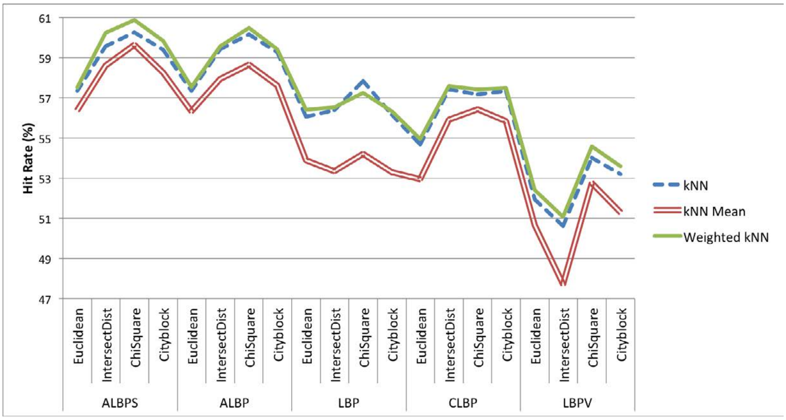


Figure 5.1: Results achieved on the KTH tips2-a dataset using different classifiers. The hit rate value corresponds with the mean hit rate along different values of k (1, 3, 5, 7, 9, 11) for all the pairs Descriptor-Distance.

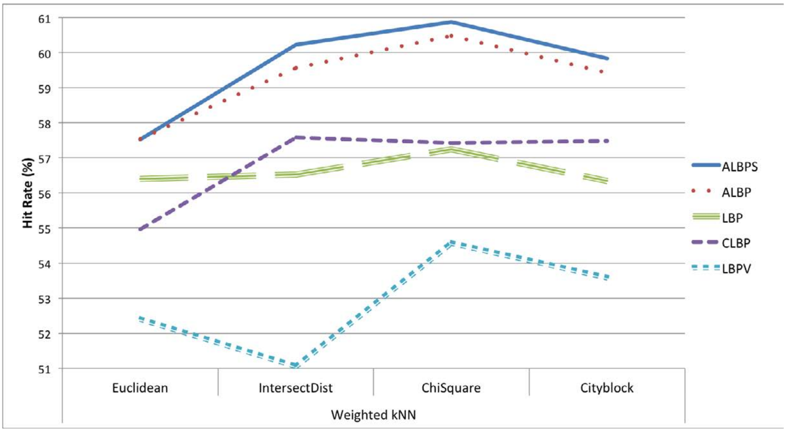


Figure 5.2: Results achieved on the KTH tips2-a dataset using different descriptors and the weighted kNN classifier. The hit rate value corresponds with the mean hit rate along different values of k (1, 3, 5, 7, 9, 11) for each distance.

Best k Neighbour Value

As it has been proven, the best result was achieved using Chi square and ALBPS. However, this result is the mean value along different numbers of neighbours. In Figure 5.3, we can see the performance of ALBPS for different k values and all the distances evaluated. The results show that the best hit rate is obtained with Chi square for k=1 and equals 61.47%.

In Table 5.1, we can see the results achieved using the weighted kNN metric for

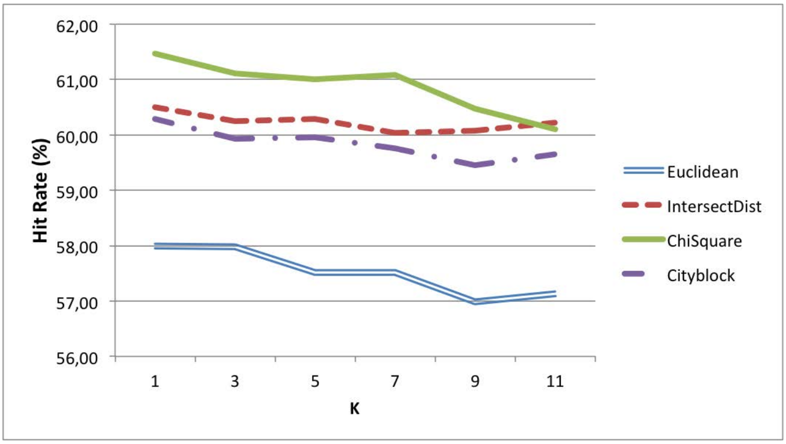


Figure 5.3: Results achieved on the KTH tips2-a dataset using weighted kNN and different values of k and distance metrics.

all the distances highlighting the best k result.

Table 5.1: Hit rate in % using weighted kNN with several distances on the KTH tips2-a dataset.

Classifier method		ALBPS	ALBP	LBP	CLBP	LBPV
kNN weighted	Euclidean	58.00	58.00	56.90	55.18	52.97
	Intersect	60.50	60.06	56.88	58.31	51.47
	ChiSquare	**61.47**	60.82	57.58	58.08	54.78
	Cityblock	60.29	59.85	56.57	58.12	53.79

5.3. Vitality sperm classification

5.3.1. Experimental Setup

As we did in the previous section for material recognition, three kNN approaches (standard kNN, mean distance kNN and weighted kNN) have been used to classify the spermatozoa images as dead or alive. We have used the one versus all paradigm which consists of classifying each image taking into account the rest of the images in the dataset. This process has been repeated using several values of neighbours (1, 3, 5, 7, 9 and 11).

Furthermore, we have evaluated the texture methods using a supervised classification. Once the 450 images are range filtered and the ground truth vector is obtained using the fluorescent images, a cross validation algorithm has been implemented in order to avoid biased results. First of all, a 20% random subset of the total number of images is kept back in order to get the test results. With the rest of the images, a 10-fold crossvalidation has been carried out. Classification was accomplished using Support Vector Machine (SVM) with Least Squares (LS) training algorithm and a linear kernel. Experiments using different kernels have been performed obtaining worse results than with the linear one.

Since this dataset is skewed due to the high number of alive heads in contrast with the low number of dead ones, F-Score has been used as one of the quality metrics, and it was applied over the skewed class, in this case, the dead images subset.

F-Score has been computed as $F - Score = 2 \cdot Precision \cdot Recall/(Precision + Recall)$, where Precision has been computed as $Precision = TP/(TP + FP)$ and Recall as $Recall = TP/(TP + FN)$, being TP the number of true positives in the classification, TN the number of true negatives and, FP and FN the false positives and false negatives respectively. Note that the positive class is the one with less elements in the training set, in this instance the positive class corresponds to the dead heads.

F-Score results are in the range $[0, 1]$ where values near 0 indicate a poor classification and values close to 1 show a good performance.

As the process of selecting the training and cross validation subsets is a random one, sometimes the number of images in each class could be quite balanced. For this reason, the accuracy measure has been taken into account, and computed as $Accuracy = (TP + TN)/(TP + FP + TN + FN)$.

Therefore, we obtained F-Score and accuracy measures for all 10-folds combinations of training and cross validation datasets. Afterwards, the classifier which outperformed the others both in terms of accuracy and F-Score has been selected as the best classifier and its parameters were used to classify the test set in order to get a more reliable performance. By using a cross validation set instead of directly a test set we avoid that the decision about the best classifier is influenced by the random cross validation set and the classifier cannot generalize well to future test samples.

In Figure 5.4 a scheme of this process is showed step by step.

5.3.2. Results with kNN

Nearest Neighbours variants evaluation

We compare ALBPS proposal with the original ALBP method and the base one LBP. In Figure 5.5, we can observe the performance of different configurations of ALBP and our ALBPS proposal using three distance metrics: Euclidean, Chi square and Cityblock. In this experiment we have excluded the Intersect distance due to its

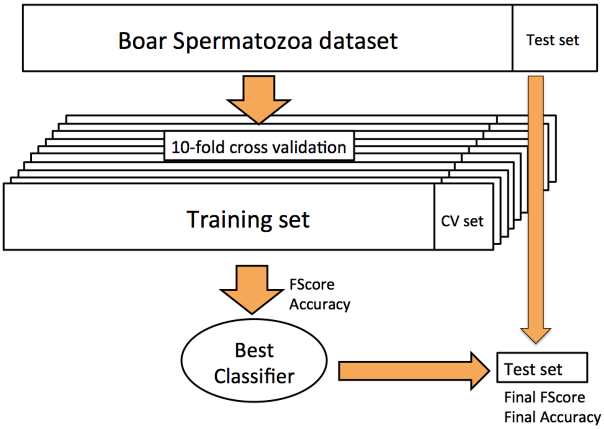


Figure 5.4: Scheme of the experimental setup.

poor performance. As in the material recognition section, the results show the mean performance along different values of k. As we can see, in the vitality classification problem, the weighted classifier does not improve the original kNN whereas the kNN with mean distances obtains lower performance in all cases. As the simple kNN has lower computational cost than the weighted kNN, we selected the former one as the best classifier.

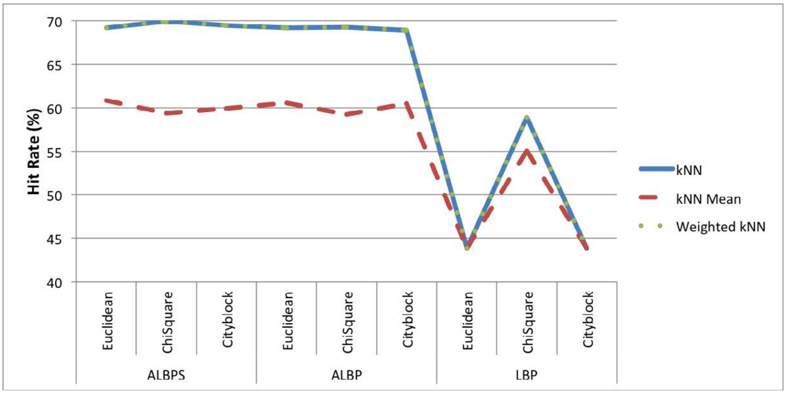


Figure 5.5: Results achieved on the boar sperm dataset using different descriptors and the simple kNN classifier. The hit rate value corresponds with the mean hit rate along different values of k (1, 3, 5, 7, 9, 11) using different distances.

LBP Variants Performance with Weighted NN

The proposed ALBPS method has been compared with the ALBP method and the original LBP in order to determine its performance. In Figure 5.6, we can see the results obtained by these descriptors using kNN and different distance metrics. ALBPS outperforms the others in all the experiments, achieving an improvement of 1.05% in hit rate with respect to ALBP and a 18,80% with respect to the original LBP method. This Figure shows that the Chi square distance, as well as in the material recognition dataset, proves to outperform the rest.

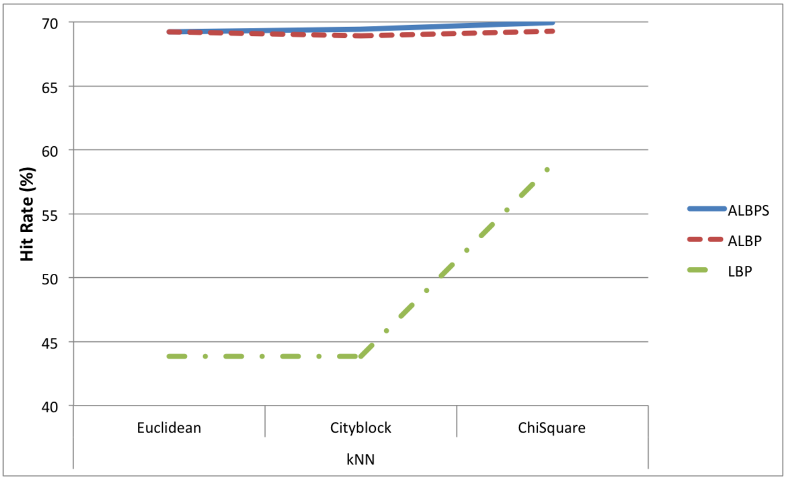


Figure 5.6: Results achieved on the boar sperm dataset using ALBP, ALBPS and the original LBP. The hit rate value corresponds with the mean hit rate along different values of k (1, 3, 5, 7, 9, 11) for all the pairs Descriptor-Distance.

Best k Neighbour Value

In order to determine the best configuration, In Figure 5.7 we show the performance of ALBPS for all the k values and all the distances evaluated. Results show that the best distance remains the Chi square with a hit rate of 72,66% for k=9.

In Table 5.2, the results obtained using kNN with all the distances choosing the best k for each descriptor are numerically presented.

5.3.3. Results with SVM

In this subsection, we show the performance evaluation results using ALBPS and other state of the art methods and SVM in terms of the F-Score and accuracy.

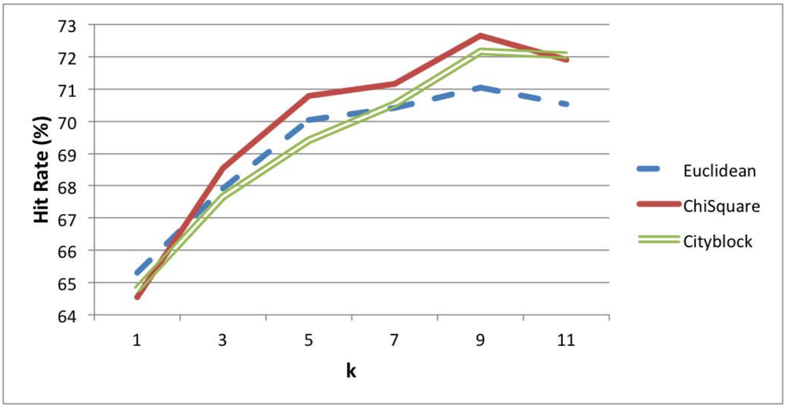


Figure 5.7: Results achieved on the boar sperm dataset using kNN with different k values and distances.

Table 5.2: Hit rate in % using kNN on the sperm dataset.

Classifier method		ALBPS	ALBP	LBP
	Euclidean	71.03	71.41	43.82
kNN	ChiSquare	**72.66**	71.79	59.30
	Cityblock	72.16	71.54	43.80

Performance evaluation using Global Texture Descriptors

In our first approach several well-known global texture descriptors have been evaluated. In Table 5.3 it is possible to see the F-Score, Precision, Recall and Accuracy achieved with these descriptors whereas in Figure 5.8 F-Score and accuracy results and how they are directly related are shown graphically.

Table 5.3: Performance of global descriptors

	WCF13	WCF4	Statistical	Legendre	Flusser	Zernike	Hu
F-Score	**0.800**	0.795	0.783	0.780	0.739	0.736	0.720
Precision(%)	**80.43**	79.54	78.72	71.00	75.58	71.38	66.24
Recall(%)	**79.57**	79.54	78.10	86.59	72.65	76.29	78.76
Accuracy(%)	**76.88**	76.75	75.69	75.00	70.88	70.19	65.44

As it can be noticed, using WCF13 the performance improved compared to the rest of global descriptors, yielding both the best F-Score value and the best accuracy. In contrast, values from Hu, Zernike and Flusser moments are quite low, Hu just obtained a 65% of accuracy which is an unacceptable result for this problem. Therefore, we can conclude that global texture descriptors offer poor results for assessing the vitality of boar semen samples.

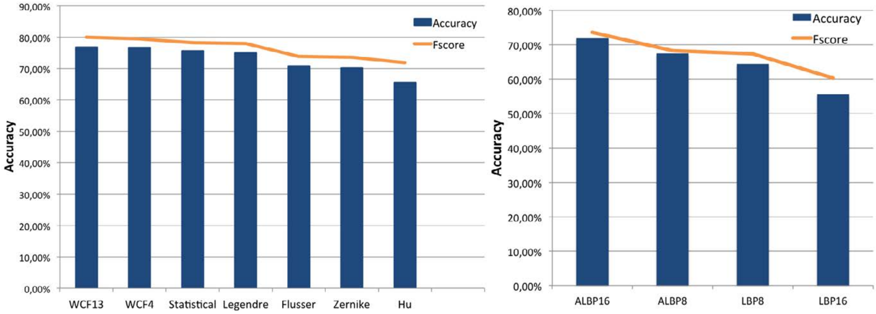


Figure 5.8: Performance of different global texture descriptors (left) and performance of local texture descriptors LBP and ALBP using $R = 1$, $P = 8$ and $R = 2$, $P = 16$ as neighbourhoods (right). The F-Score is extended to the range [0-100] in order to preserve the legibility of the graphic.

Performance evaluation using Local Texture Descriptors

In this experiment, we used two local texture descriptors: LBP and the adaptive version ALBP proposed by Guo, Zhang, Zhang and Zhang (2010).

Two different neighbourhoods, $R = 1$, $P = 8$ and $R = 2$, $P = 16$ have been used, to measure F-Score and accuracy in both cases. Their performance can be seen in Table 5.4 and Figure 5.8 (right).

Table 5.4: Performance of local texture descriptors LBP and ALBP using $R = 1$, $P = 8$ and $R = 2$, $P = 16$ as neighbourhoods.

	ALBP2,16	ALBP1,8	LBP1,8	LBP2,16
F-Score	**0.737**	0.683	0.674	0.603
Precision(%)	**73.26**	67.47	67.05	60.00
Recall(%)	**74.12**	69.14	67.82	60.67
Accuracy(%)**71.88**	67.50	64.38	55.63	

ALBP behaves better than LBP in all cases. ALBP2,16 obtains similar results to global descriptors, with a F-Score equals to 0.737 and an accuracy of 71.88%, outperforming in more than a 15% the accuracy of the classical LBP2,16. Nevertheless, global WCF13 descriptor outperforms ALBP approach.

ALBPS compared with previous Local Texture Descriptors

In this experiment, the performance using our proposed method, ALBPS, which includes oriented standard deviation information for the images description has been assessed. To measure the performance of our proposal, we compared it with LBP and ALBP methods. Results when concatenating LBP histogram from the original LBP method with oriented standard deviation vector (LBPS) have also been obtained.

In Table 5.5 and Figure 5.9, a comparison between ALBPS and the original ALBP and LBP methods is shown. Adding oriented standard deviation to local texture descriptors, LBPS and ALBPS, improves both original LBP and ALBP methods which verifies the effectiveness of our proposal. The best overall result is achieved with ALBPS with $R = 2$, $P = 16$ (ALBPS2,16) yielding a F-Score equals to 0.842 and a 81.88% of accuracy which means an improvement of 14.25% in F-Score and of 13.91% in accuracy over the base method, ALBP2,16. Moreover, we would like to highlight that ALBPS2,16 also outperforms global texture description, specifically, by a 5.25% in F-Score and by a 6.5% in accuracy with regard to WCF13. It is also noticeable that ALBPS2,16 also outperforms the previous related works found in the literature since the best approach (Garcia-Olalla et al., 2012) obtained a hit rate of just 78.67%. Therefore, in the carried out experiments is clear that our proposed method outperforms global traditional descriptors, previous local texture descriptors based on LBP and previous related works.

Table 5.5: Performance of our proposed texture descriptors compared with the best of previous experiments ALBP2,16.

	F-Score	Precision (%)	Recall (%)	Accuracy (%)
ALBPS2,16	**0.842**	**83.70**	**84.62**	**81.88**
ALBPS1,8	0.753	76.50	74.27	72.13
LBPS2,16	0.747	73.61	76.17	70.50
ALBP2,16	0.737	73.26	74.12	71.88
LBPS1,8	0.710	71.07	71.10	69.25

A new improvement: combining Global and Local features

Our last experiment consisted of combining the analysed local descriptors with global features into a new hybrid features vector. We intended to introduce global context to resolve ambiguities that can occur locally when an image has multiple similar regions. Consequently, WCF13 and WCF4 were merged with the studied local descriptors yielding the results shown in Table 5.6 and Figure 5.10. The best overall result was achieved when combining WCF13 and ALBPS2,16 reaching a F-Score of 0.886 and a 85.63% of accuracy outperforming the results obtained with

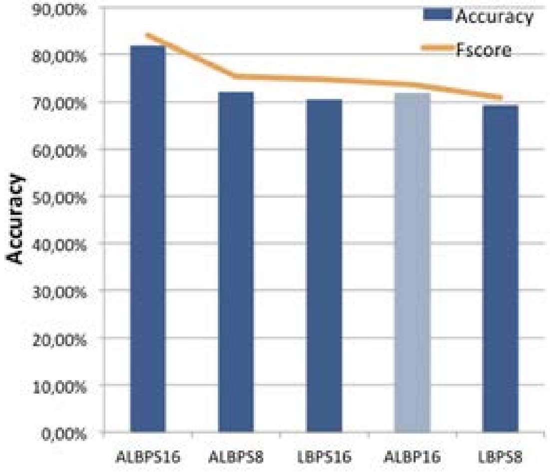


Figure 5.9: Performance of our proposed texture descriptors and the best of previous local texture descriptors ALBP2,16 (dotted bar). The F-Score is extended to the range [0-100] in order to preserve the legibility of the graphic.

local and global texture separately. Particularly, WCF13+ALBPS2,16 improves F-Score value in a 5.23% and accuracy in a 4.58% with regard to our individual local descriptor ALBPS2,16.

Figure 5.10: Performance of hybrid global+local texture descriptors. The F-Score is extended to the range [0-100] in order to preserve the legibility of the graphic.

It is important to note the high value of recall (89%) obtained, which means that the algorithm detects a high percentage of dead heads. Specifically, it only misclassifies a 11% of them as alive ones. Since there are many more images of alive heads than dead ones in our dataset, predicting that a new test image is alive has a higher probability of chance than otherwise. A value of recall equals to zero would imply that the algorithm is just classifying all images as alive (no skewed class) and therefore it is obtaining a high accuracy next to 100% without being useful for the goal of our task. As a consequence, our approach is correctly classifying the database even though the existence of a skewed class.

Table 5.6: Performance of hybrid global+local texture descriptors.

	F-Score	Precision (%)	Recall (%)	Accuracy (%)
WCF13+ALBPS2,16	**0.886**	**88.12**	**89.00**	**85.63**
WCF13+LBPS1,8	0.867	89.77	84.04	85.00
WCF13+LBP1,8	0.865	93.67	80.43	85.63
WCF4+LBPS2,16	0.865	87.50	85.56	85.00
WCF13+ALBPS1,8	0.860	84.21	87.91	83.75
WCF13+LBPS2,16	0.857	80.41	91.76	83.75
WCF13+ALBP2,16	0.854	86.42	84.34	85.00
WCF4+ALBPS1,8	0.852	87.80	82.76	84.38
WCF13+ALBP1,8	0.850	85.87	84.04	82.50
WCF4+LBP1,8	0.847	86.02	83.33	81.88
WCF13+LBP2,16	0.845	82.83	86.32	81.25
WCF4+ALBPS2,16	0.843	84.27	84.27	82.50
WCF4+LBPS1,8	0.832	81.19	85.42	79.38
WCF4+ALBP2,16	0.825	82.02	82.95	80.63
WCF4+ALBP1,8	0.818	76.29	88.10	79.38
WCF4+LBP2,16	0.806	79.80	81.44	76.25

Besides, we want to highlight that this hybrid WCF13+ALBPS2,16 descriptor classified with SVM obtains better results than the previous works in the literature where we found hit rates of 76.80% in Alegre, Garcia-Olalla, Gonzalez-Castro and Joshi (2011); Alegre, Garcia-Ordas, Gonzalez-Castro and Karthikeyan (2011) and 78.67% in Garcia-Olalla et al. (2012). Therefore, the approach presented in this thesis obtains about a 8.85% of hit rate improvement over previous works.

5.4. Conclusions

In this chapter, we proposed a new local texture descriptor ALBPS by adding an oriented standard deviation term to the ALBP descriptor. It also has been proved that adding this new term to the classical LBP its performance also increases.

We have evaluated that method in two different application fields: material recognition and Vitality sperm classification.

In the first experiment, we have demonstrated that the proposed method ALBPS outperforms recent LBP descriptors variants, such as ALBP, CLBP and LBPV as well as the classical LBP method for several kNN-based methods with multiple distances on the multi class KTH-tips 2a dataset for material recognition. Results show that the best descriptor was ALBPS using a weighted kNN and the Chi square distance achieving a 61.47% of hit rate.

In the second one, we have combined the local proposed descriptor, ALBPS2,16 with the global WCF13 descriptor obtaining a features vector which contains local and global information. The experimental results showed that the hybrid features extracted by the proposed method provide a better performance than previous works when using a robust SVM classification. Also, we were able to ascertain that the skewed class, the dead one, was successfully classified reaching a recall of 89%. A F-Score equals to 0.886 and an accuracy of 85.63% were yielded by the early fusion of WCF13 and ALBPS2,16 which is a very interesting result for classifying the vitality of boar spermatozoa heads as dead or alive. Furthermore, we have also demonstrated the good performance of ALBPS using several kNN-based methods with multiples distances against LBP descriptors variants such as ALBP, CLBP and LBPV as well as the classical LBP method. ALBPS proved to outperform the rest of the methods reaching a hit rate of 72.66%. However the weighted kNN does not show any improvement with respect to the standard kNN, being the simple kNN method with k=9 and Chi square distance the best classifier.

Taking into account all the results obtained, we can conclude that adding standard deviation information to the feature descriptor, as proposed with ALBPS, the accuracy was incremented on both datasets with SVM and all combinations of evaluated non parametric classifiers and metrics which gives a positive insight on the use of ALBPS in very different scenarios.

Chapter 6

Local Oriented Statistical Information Booster (LOSIB)

6.1. Methodology

The main purpose of the Local Oriented Statistical Information Booster (LOSIB) is to enhance the performance of a texture descriptor.

The basic concept is to add local oriented statistical information computed along all pixels of the image. This information is rarely taken into account when texture is described and gives extremely useful information for texture discrimination. In this work, the combination of LOSIB with widely used texture descriptors was done by means of the concatenation of both vectors.

As the depth of the neighbourhood increases, LOSIB comprises information at a more global scale. Depending on the image dataset, local or more loose global information can achieve better performance.

Another factor is the number of neighbours in the neighbourhood. In this sense, more neighbours means that a higher number of different orientations have been taken into account. As the texture becomes more heterogeneous, the number of neighbours should be increased in order to capture all the variety of the image. However, using excessive orientations on homogeneous textures may be counterproductive due to the loss of weight of the important ones.

Therefore, the nomenclature for this method is LOSIB(R,P) where R is the radius of the neighbourhood and P the number of neighbours.

Let c be a pixel at position (x_c, y_c) of the image, p be a pixel of its neighbourhood (with $p \in \{0, 1, ..., (P-1)\}$), whose coordinates are (x_p, y_p), and let g_c and g_p be their respective grey level values. In order to obtain the LOSIB of an image, it is first necessary to extract the absolute differences d_p between the grey level values g_c and g_p, for all pixels c of the image, as shown in Equation (6.1). Figure 6.1 depicts an example of this oriented difference extraction at three pixels.

$$d_p(x_c, y_c) = |g_c - g_p| \tag{6.1}$$

Given a pixel c, the coordinates (x_p, y_p) of its p-th neighbour are obtained by

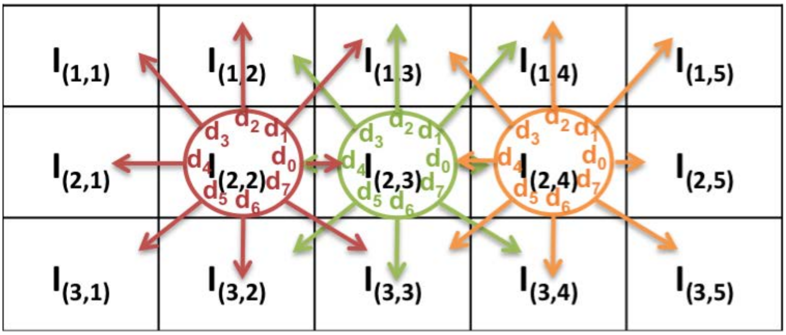


Figure 6.1: Extraction of the absolute difference of gray-level values for three pixels to compute LOSIB(1,8).

means of Equation (6.2).

$$(x_p, y_p) = (x_c + R\cos(2\pi p/P),\ y_c + R\sin(2\pi p/P)) \tag{6.2}$$

The values of the neighbours that are not in the centre of grids can be estimated by interpolation of their connected pixels.

Then, the mean of all the differences along the same orientation is computed following Equation (6.3), where N and M are the number of rows and columns of the image, respectively.

$$\mu_p = \frac{\sum_{x_c=1}^{M} \sum_{y_c=1}^{N} d_p(x_c, y_c)}{M \cdot N} \tag{6.3}$$

In Figure 6.2 the histogram of all the absolute differences along the orientation $p = 0$ and the value μ_0 for an image of KTH-TIPS 2a dataset is shown.

Thereby, LOSIB will have as many features as neighbours are in the considered neighbourhood and it represents the mean difference for all the orientations shown in Equation (6.4).

$$\text{LOSIB}(R, P) = \bigcup_{p=0}^{P-1} \mu_p \tag{6.4}$$

An example of the orientations of LOSIB(1,8) is shown in Figure 6.3 for clarification.

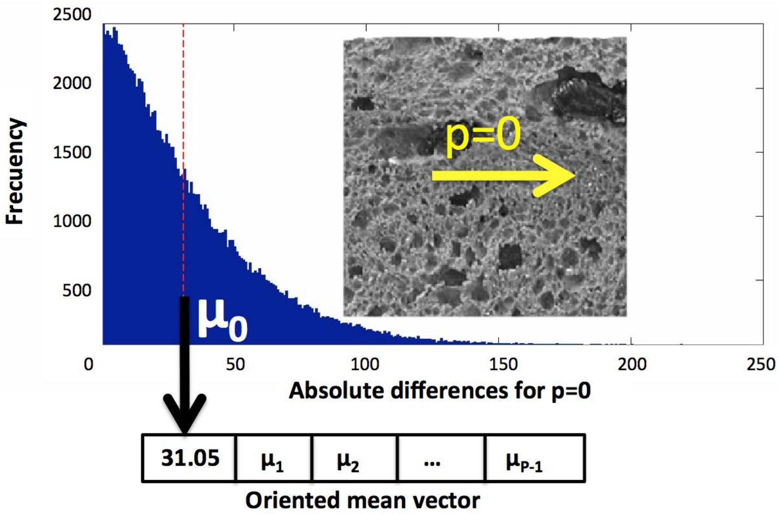


Figure 6.2: Example of a histogram of all the absolutes differences along the orientation 0° (thus, $p = 0$) and the mean value used to yield the final LOSIB.

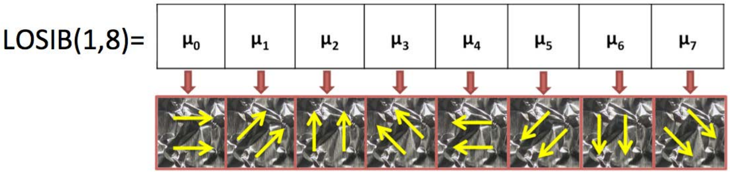


Figure 6.3: Different orientations using 8 neighbours.

6.2. Material recognition

6.2.1. Experimental Setup

KTH TIPS2-a dataset

The experimental setup used for the KTH TIPS2-a dataset is the standard protocol developed by Caputo et al and used in several works (Caputo et al., 2005; Chen, Shan, He, Zhao, Pietikainen, Chen and Gao, 2010) as explained in Chapter 5. Again, we carried out four classifications using this method to increase its robustness, one classification using each texture sample as the test set. The mean of the hit rate in each iteration was computed. We have used a Support Vector Machine (SVM) with the one-vs-one paradigm to classify the images. We have selected the Least Squares training algorithm and a polynomial kernel of order 2 after some performance tests.

Furhermore, we have also classified with a non parametric k-Nearest Neighbour (kNN) algorithm in the same way as we did in the previous chapter.

Brodatz32 dataset

The setup of the Brodatz32 experiments is quite similar to the KTH TIPS2-a one. However, Brodatz32 does not have different samples in each class so we have used a 4-fold cross validation in order to avoid randomness, extracting the average as the final accuracy result. In this way, we have used a 75% of images as training set and the remaining 25% as the test set. A SVM trained with Least Squares and a polynomial kernel with order 2 to find the decision boundary has been used. Since the dataset is multiclass, we have chosen again the one-vs-one paradigm.

6.2.2. KTH TIPS2-a dataset results

The first experiment deals with the classification of KTH TIPS2-a using a one-vs-one SVM paradigm and it demonstrates that the LOSIB enhancer improves the hit rate over all the classical descriptors. Three configurations of LOSIB have been evaluated: LOSIB(1,8), LOSIB(2,16), and a concatenation of both (which will be called LOSIB(1,8)+LOSIB(2,16)). Figure 6.4 depicts the results using classical global descriptors. It is specially remarkable the cases of the Hu, Legendre, Flusser and Zernike moments. They yield a poor performance by themselves alone, but LOSIB makes their accuracy to increase more than a 50% in all cases (161% in the case of the Hu moments) The best hit rate is 63.33%, obtained by Haralick and LOSIB(1,8)+LOSIB(2,16). The improvement of the most recently used global descriptors is clearer shown in Figure 6.5. Methods based on the Haralick features, computed from the co-occurrence matrix of the texture directly (called Haralick), or from the Wavelet response (called WCF4 and WCF13 (Alegre et al., 2012)) obtain the best results in combination with with the LOSIB(1,8)+LOSIB(1,16), outperforming the descriptor in a 16.81% with WCF4. However, using the Wavelet Statistical Features (WSF) the best improvement was obtained in combination with just LOSIB(2,16) (6,91%).

Figure 6.5 shows the results using descriptors and their combination with LOSIB. Even though the amelioration of their early fusion with LOSIB is not high as in the case of global descriptors, it is evident that LOSIB provides useful information that improves the classification results.

In Table 6.1 the hit rate obtained in all the tests carried out is shown. The best hit rate was 71.44%, achieved with CLBP and LOSIB(1,8) but the higher improvement was obtained with LOSIB(1,8) and LBPV (8.28% of hit rate improvement). LBP

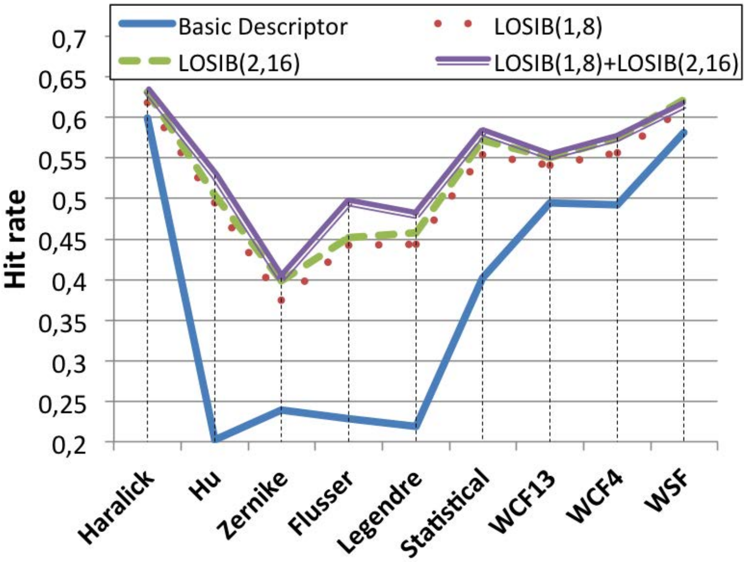


Figure 6.4: Results on the KTH-TIPS 2a dataset using the global descriptors and the combination with the LOSIB(1,8), the LOSIB(2,16) and both of them.

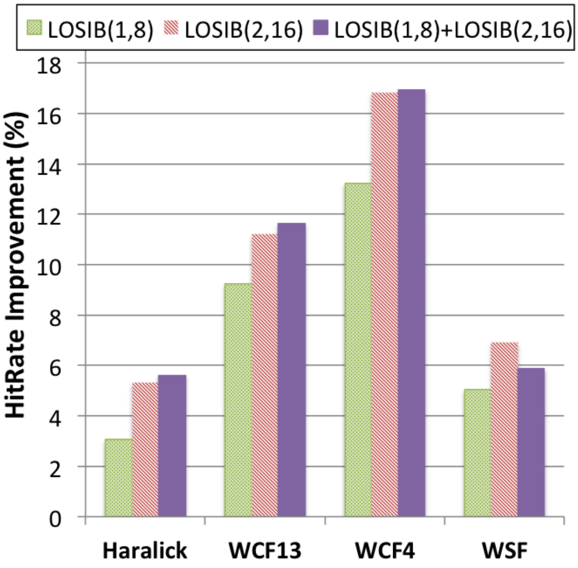


Figure 6.5: Improvement in terms of hit rate of commonly used global descriptors in the last years when they are combined with LOSIB.

descriptor achieved lower hit rates when it is combined with LOSIB(1,8) (65.11%) and LOSIB(2,16) (65.17%), rather than by itself alone (65.53%), but obtained better results with the concatenation of both of them (66.83%).

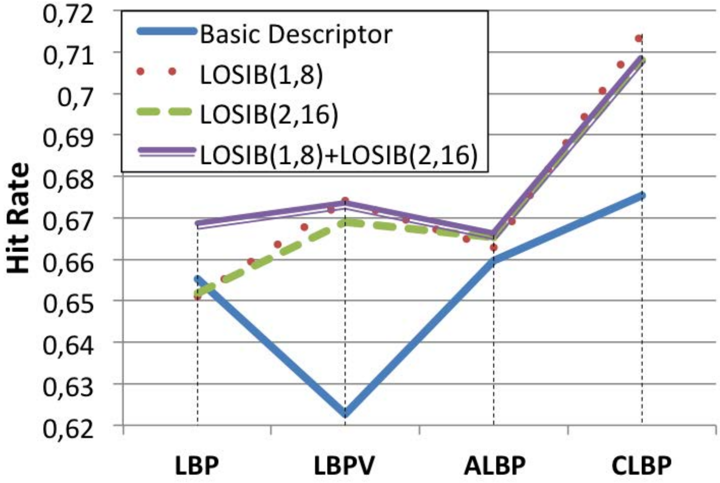


Figure 6.6: Results on the KTH-TIPS 2a dataset using the LBP descriptors and the LOSIB(1,8), LOSIB(2,16) and LOSIB(1,8)+LOSIB(2,16).

In Figure 6.7 we can clearly see the improvement that LOSIB introduces in combination with the LBP-based descriptors. As already mentioned, LBP descriptors are just outperformed by their combination with LOSIB(1,8)+LOSIB(2,16), but the hit rate of all the others is increased for the three configuration of LOSIB. The higher improvement in terms of hit rate reaches a 8.28% using LOSIB(1,8) and LBPV.

The good performance of LOSIB enhancer is confirmed by the experiments carried out using the kNN classifier. Once again, the hit rate of all descriptors is increased when they are combined with the three boosters assessed (LOSIB(1,8), LOSIB(2,16) and LOSIB(1,8)+LOSIB(2,16). In Figure 6.8 we can see the results obtained with kNN using the Chi square distance metric. The highest improvement in terms of hit rate was achieved with the global descriptors while the best result was obtained using ALBP with a hit rate of 62.10% when combining it with LOSIB(1,8)+LOSIB(2,16).

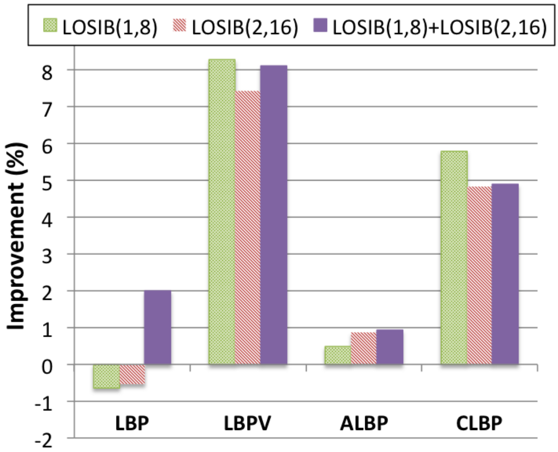


Figure 6.7: Improvement in terms of hit rate of commonly used LBP-based local descriptors in the last years. when they are combined with LOSIB.

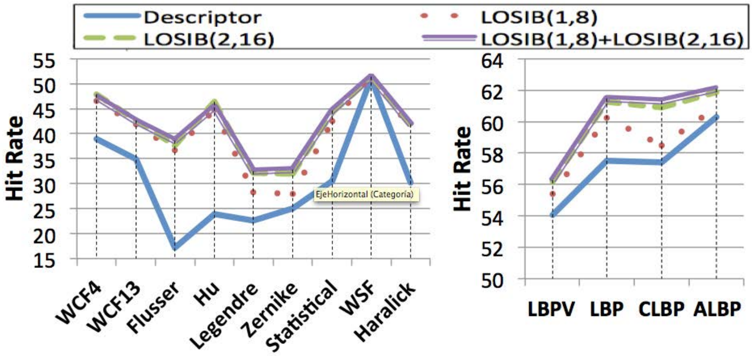


Figure 6.8: Results using kNN on the KTH-TIPS 2a dataset. In the left, the global descriptors and in the right the local LBP family.

6.2.3. Brodatz32 dataset results

In order to assess the robustness of LOSIB in material recognition tasks, more tests have been carried out using Brodatz32 dataset. In the same way, a one-vs-one SVM have been used but in this case with a 4-fold strategy. The results are shown in Figure 6.9. In all cases, LOSIB booster enhances the results of the descriptors, being the best hit rate 91.06% with LBPV and both LOSIB configurations. The higher improvement in terms of hit rate equals to 52.29% was obtained again combining Haralick descriptor with LOSIB(1,8)+LOSIB(2,16). In Table 6.2 the numerical results

Table 6.1: Performance, in %, of the combinations of LOSIB with the global and local descriptors on the KTH-tips 2a dataset when classifying with SVM.

Descriptor	Haralick	Hu	Zernike	Flusser	Legendre	Statistical	WCF13
Alone	59.95±2.57	20.22±2.97	23.88±1.54	22.87±5.91	21.86±4.76	40.13±5.06	49.56±4.23
LOSIB(1,8)	61.78±5.51	49.49±5.67	37.42±4.19	44.19±5.37	44.30±7.03	55.37±6.98	54.08±3.76
LOSIB(2,16)	63.15±4.37	50.40±4.97	39.84±3.34	45.16±4.36	45.79±6.05	57.15±6.18	55.07±4.10
LOSIB(1,8)+LOSIB(2,16)	**63.32**±2.96	**52.90**±4.16	**40.13**±4.13	**49.60**±4.14	**48.02**±6.75	**58.22**±3.31	**55.28**±3.37

Descriptor	WCF4	WSF	LBP	ALBP	CLBP	LBPV
Alone	49.18%±4.39	58.14±2.98	65.53±3.61	65.97±5.63	67.53±3.17	62.27±5.81
LOSIB(1,8)	55.68%±6.66	61.07±4.31	65.11±7.98	66.29±7.98	**71.44**±5.16	**67.42**±8.21
LOSIB(2,16)	57.45%±6.62	**62.16**±4.79	65.17±8.32	66.54±8.24	70.79±5.50	66.90±8.38
LOSIB(1,8)+LOSIB(2,16)	**57.51**%±5.74	61.55±4.25	**66.84**±8.00	**66.58**±8.56	70.83±5.59	67.32±6.73

are summarised.

Table 6.2: Performance, in %, of the combinations of LOSIB with the global and local descriptors on the Brodatz32 dataset when classifying with SVM.

Descriptor	Haralick	WSF	WCF13	WCF4
Alone	56.59±5.45	65.47±3.66	76.76±2.80	82.81±1.77
LOSIB(1,8)	84.13±1.42	71.58±3.21	80.03±3.15	87.70±2.29
LOSIB(2,16)	85.64±2.08	74.27±3.03	83.44±2.85	88.28±3.38
LOSIB(1,8)+LOSIB(2,16)	**89.01**±2.20	**77.15**±3.21	**83.94**±2.45	**89.06**±3.00

Descriptor	LBP	ALBP	LBPV	CLBP
Alone	74.90±2.51	79.39±3.48	85.10±3.25	81.88±2.23
LOSIB(1,8)	85.89±3.87	86.18±3.43	87.65±3.19	83.84±2.11
LOSIB(2,16)	87.35±2.43	88.18±3.51	90.28±3.13	**87.70**±1.66
LOSIB(1,8)+LOSIB(2,16)	**88.18**±2.58	**88.57**±2.29	**91.06**±4.04	87.55±2.03

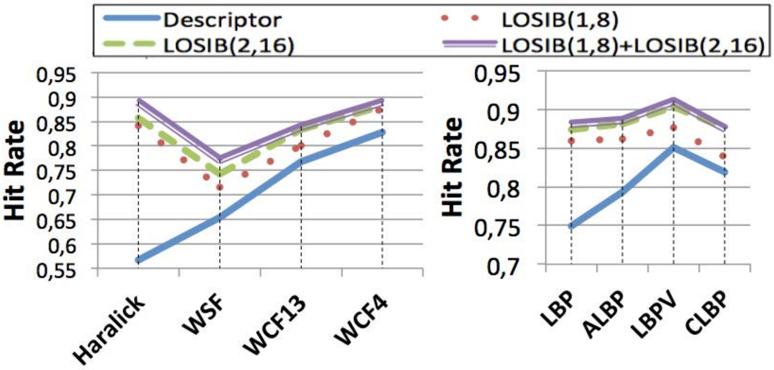


Figure 6.9: Results using SVM on the Brodatz32 dataset. In the first table, the global descriptors and in the second, the local LBP-based family.

6.3. Acrosome integrity classification

6.3.1. Experimental setup

Several configurations of the classifier have been tested in order that our statistical analysis is able to generalize to an independent data set. A cross validation algorithm has been implemented to deal with this model selection problem. Therefore, the parameters of the classifier will be chosen testing the cross validation set and once the best parameters are selected, the results will arise using the test set. To that end, a random subset of 20% of the images are kept to use as test set. Since this is a dichotomous classification, the splits of our dataset will be done in a way that each subset contains roughly the same proportions of the two classes, intact and damaged. Then, a 10-fold cross validation algorithm is applied to choose the best parameters and solve the problem of the model selection. The classification was performed with a Support Vector Machine (SVM) backed by a Least Squares (LS) training algorithm and a linear kernel. Experiments with Gaussian kernel obtained worst results. This process is completely explained in the algorithm 1 and a schema is shown in Figure 6.10.

Algorithm 1 Cross validation algorithm and experimental setup.

```
Require: Cropped heads.
Ensure: Precision, recall, F-Score and accuracy of the method.
    Select randomly 20% of the images as test set (20% of the intact images and 20%
    of the damaged images).
    Split the rest of the images (80%) in 10 subsets with equally intact and damaged
    heads proportions.
    for each parameter configuration do
        for i=1:10 do
            Take the $i^{st}$ subset as cross validation set.
            Take the rest as training set.
            Perform SVM with LS training.
            Compute and store F-Score and accuracy.
        end for
        Compute and store mean F-Score and accuracy on the ten subsets.
    end for
    Select the SVM model with the highest F-Score and accuracy.
    Evaluate test set using the selected SVM.
    Compute precision, recall, F-Score and accuracy.
```

Precision, recall, F-Score and accuracy are the measurements used to assess the response of our method. The main reason is because they are well known and widely employed by the pattern recognition community. Besides, these metrics were used in previous works therefore it is easier to compare the results obtained in

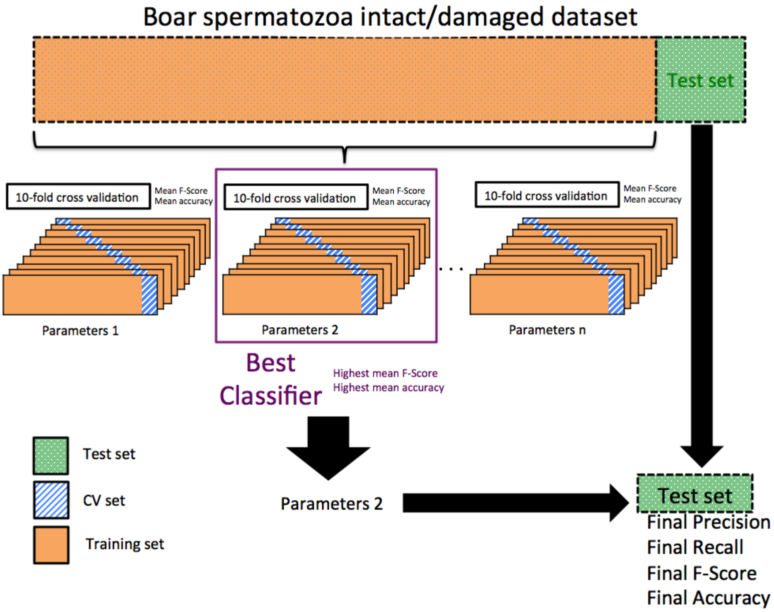


Figure 6.10: Scheme of our cross validation experimental setup.

this work with the state of the art. We defined the damaged heads as the positive class and the intact one as the negative class. Then, a True Positive (TP) is a damaged head classified as damaged; a True Negative (TN) is an intact head recognized as intact; a False Positive (FP) is an intact head classified as damaged; and a False Negative (FN) is a damaged head recognized as intact. Thereafter, precision, recall, F-Score and accuracy are defined in Equations (6.5), (6.6), (6.7) and (6.8) respectively. Note that F-Score varies in the range $[0, 1]$ being 1 the best score and 0 the worst one.

$$precision = \frac{TP}{TP + FP} \tag{6.5}$$

$$recall = \frac{TP}{TP + FN} \tag{6.6}$$

$$F - Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{6.7}$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{6.8}$$

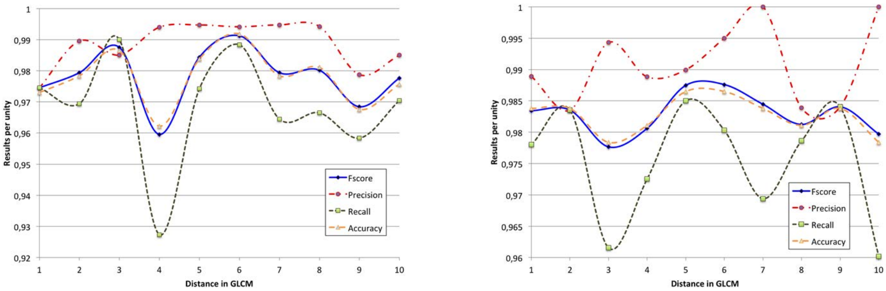
6.3.2. Results

In Figure 6.11, the results for each evaluated descriptor are presented. As mentioned, the WCF13 (W) is evaluated for distances of the GLCM from 1 to 10 whereas the LBP approach is shown both for radius 1 and 8 neighbours (L8) and for radius 2 and 16 neighbours (L16). Fourier contour descriptor will be shorten as F. The considered descriptors are: the basic approach of only considering WCF13 descriptor, the combination of WCF13 either with Fourier (WF) or with LBP (WL8 and WL16) and, finally, the fusion of the three methods (WFLP8 and WFLP16). It is possible to notice that using only the global descriptor, WCF13, W in Figure 6.11 *(f)* , the best recall, and at the same time the worst precision, is obtained at distance 7. It means that almost all the damaged heads are classified rightly but also a not small number of intact ones have been considered as damaged. With this descriptor, the most balanced point is with distance 2, where the FScore is the highest because there are less intact spermatozoa classified as damaged. Therefore, this global descriptor works better when considering the information that is at short distance, classifying well the damaged heads, 99%, but also miss-classifying 4 intact heads out of 100. Hence, global descriptors provide us a high recall.

When adding local information, coming from LBP, Subfigures *(d)* and *(e)* results change slightly. The information of the 8 neighbourhoods combined with Wavelet (Subfigure *(e)*) makes differences between precision and recall much smaller in general, increasing the FScore when GLCM is computed with a distance of 2 pixels. It is clear that the local information of LBP8, WL8 in Subfigure *(e)*, gives more precision to the classification so, in this case, the description is more reliable and when it classifies a head as acrosome-damaged, it is true in almost 99% of the cases. When the neighbourhood is bigger, WL16, results are worse. Recall is quite low which indicates that some of the damaged ones are considered as intact. Hence, adding local information we obtain higher precision but we loose a little bit of recall.

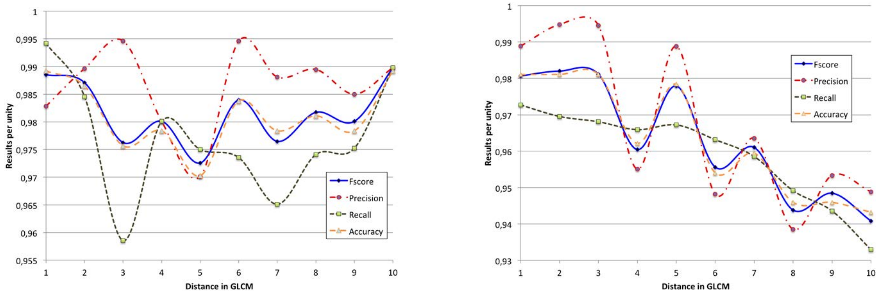
If we consider now Subfigure *(c)*, we will see that adding information of the contour, the Fourier descriptors, to the global descriptor yields a higher precision than before. Of course, depending on the distance the recall drops and again the differences between precision and recall grow. It is clear that information of the contour helps a lot for detecting the acrosome-damaged heads. Looking at Figures 4.7 and 4.6 it is noticeable that the borders of the apical area are much smoother in the intact than in the damaged ones and the Fourier descriptor is benefiting from this property for the precision detection of the acrosome-damaged heads. Therefore, adding contour information increases the precision.

Finally, in the Subfigures *(a)* and *(b)* it is possible to see the combination of the three different kind of descriptors: global (WCF13), local LBP and contour (FSD). It is remarkable that with WFLP16 a 100% of precision is reached for distances 7 and 10. The problem of this approach is that when precision is so high, recall drops



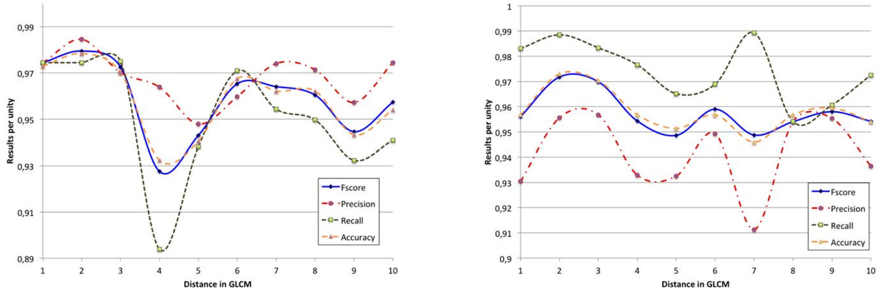
(a) WFLP8

(b) WFLP16



(c) WF

(d) WL16



(e) WL8

(f) W

Figure 6.11: Results achieved with each evaluated descriptor.

significantly. On the contrary, the same combination but using LBP with a small neighbourhood, WFLP8, yields a more balanced response, with very high Fscore at distance 3, and the highest, above 99% with distance 6. In this case, the high precision produced by the Fourier descriptors is combined with the good recall yielded by LBP8 and, in general with the appropriate behaviour of the wavelet descriptors.

By adding LOSIB information to these descriptors the performance increases in almost all the cases. The early fusion of WFLP8 and LOSIB(1,8) enhancer decreases the FScore but slightly increases the recall. However, the fusion of WFLP16 with LOSIB(2,16) achieves the best results so far for the acrosome integrity classification with an FScore at distance 1 of 0.9975, a Precision of 100% and a recall of 98.84%. Furthermore, the accuracy metric also achieved the best value of a 99.73%. In the Figure 6.12 we can see the graphic information of this method for all the different distances of the GLCM. We can conclude that adding LOSIB information to the descriptor improves the performance significantly.

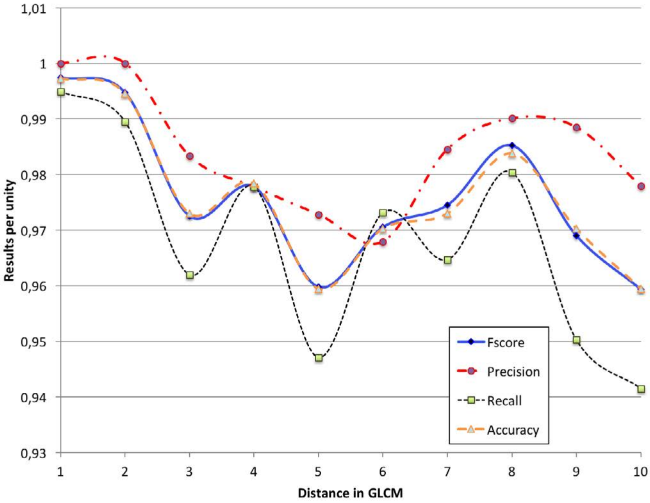


Figure 6.12: Results achieved with WFLP16 + LOSIB(2,16) enhancer.

In order to better understand these results, Figure 6.13 and Table 6.3 summarize comparatively the evaluation metrics obtained with the best distances of the GLCM for the more representative methods, including LOSIB improvement.

Figure 6.13 shows that the Wavelet descriptor alone achieves a decent recall. This means that by using global information most of the damaged heads are recognized but wrongly detecting some intact heads as damaged.

Adding Fourier contour information to the basic WCF13 descriptor (F-Score equals 0.9897 and accuracy of 98.92%) outperforms the combination of WCF13 with LBP (0.9795 F-Score and 97.84% accuracy). However both approaches perform better than the basic WCF13 (0.9718 F-Score and 97.30% accuracy). Moreover, it is interesting to notice that using global descriptors (WCF13) leads to higher recall than precision but when it is combined with local information such as LBP or Fourier the precision increases while the recall remains or even decreases. While local description can better describe the heterogeneous texture of a damage acrosome with black

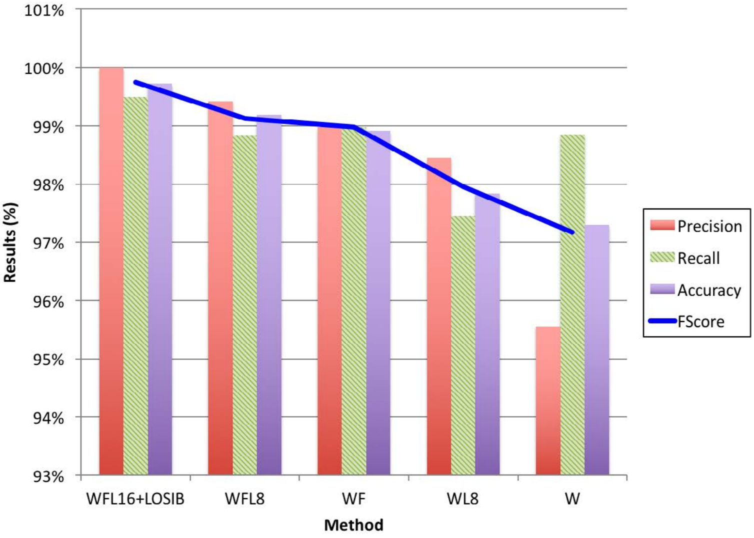


Figure 6.13: Performance of the evaluated methods with the best parameters.

Table 6.3: Results obtained for the evaluated descriptors showing just the best distances (d) in the GLCM (Gray Level Co-occurrence Matrix).

	FScore	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
WFLP16+LOSIB d=1	**0.9975**	**100.00**	**99.49**	**99.73**
WFLP8 d=6	0.9912	99.42	98.84	99.19
WF d=10	0.9897	98.98	98.98	98.92
WFLP16 d=6	0.9876	99.50	98.03	98.65
WL8 d=2	0.9795	98.45	97.45	97.84
WL16 d=5	0.9779	98.88	96.72	97.84
W d=2	0.9718	95.56	98.85	97.30

spots, global information captures better the homogeneous and smooth texture of the intact heads. Therefore the combination of both local and global methods in the description of acrosomes combined with the LOSIB enhancer is a nice solution, achieving an almost perfect performance both in terms of precision and recall.

Another important thing to highlight is that our proposal improves all previous methods in the literature dealing with the same purpose as it can be seen in Table 6.4. In the most recent one, Alegre et al. (2013) achieved a hit rate of 99% but using a very small dataset of 360 heads, 210 intact and 150 damaged. Here we have not only increased the performance but also results are more reliable since the dataset used is bigger with 1851 images (905 intact and 946 damaged).

Table 6.4: State of the art results in spermatozoa classification based on the acrosome integrity state.

	Accuracy (%)	Dataset size (intact-damaged)
N-contours and RLVQ (Alegre et al., 2013)	99	360 (210-150)
WCF4 and NN (Alegre et al., 2012)	94.93	800 (400-400)
Contour gradient and LVQ (Alegre et al., 2008)	93.2	320 (176-144)

6.4. Tool wear estimation

6.4.1. Experimental Setup

Cutting edge and Edge wear datasets

To avoid biased results due to the random factor, a cross validation algorithm was implemented using kFold. First, all the descriptors were randomly divided into ten subsets, assigning to each descriptor an index between 1 and 10. Then, four Support Vector Machine (SVM) models were trained with Least Squares (LS) training algorithm. Nine of the subsets were used for training whereas the last subset was reserved to test the classifier. At the end, the mean value of all the evaluation metrics over all iterations is shown as final result.

Wear region dataset

In the same way, with the wear region dataset the fast supervised classifier Support Vector Machine was chosen in order to learn a model which can distinguish between two or three levels of wear. In both cases, a SVM with Least Squares training algorithm and a linear kernel was used. We tested several kernels and the linear one achieved the best results. In this case, we decide to use a bigger test dataset than just the 10% used in the Cutting edge and Edge wear experiments due to the lower number of images. For that reason, a cross validation was carried out using random samples of the dataset for training the classifier (70%) and the rest for testing (30%). This evaluation was run 10 times in order to avoid random results. As we have done in previous experiments, we present as the final results the mean value of all the iterations.

6.4.2. Cutting edge and Edge wear datasets results

This section analyses the results for a classification of three levels of wear: low-medium-high (L-M-H). The best results were obtained using ALBP with LOSIB, both of them with 8 neighbours and radius 1. For this case the hit rate was 53.77% using the Edge Wear dataset. Although these results are quite low, they are helpful to demonstrate the advantages of using LOSIB booster, which improves the original results in all the evaluated cases. Results are shown graphically in Figure 6.14. Moreover, the Edge wear dataset obtains in all cases better results than the Cutting edge dataset due to the focus in the edge wear region. Another important fact is that the methods with 8 neighbours and radius 1 outperformed those with 16 neighbours and radius 2 due to the narrowness of the edge wear and the prevalence of homogeneous areas in the insert.

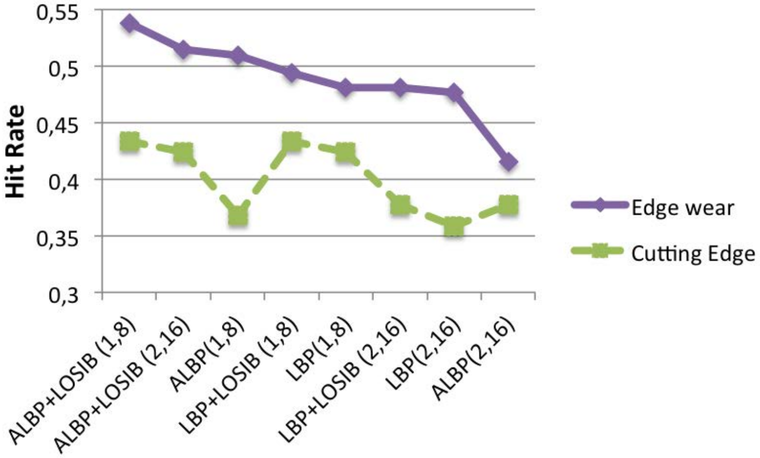


Figure 6.14: Results using low-medium-high classification on Edge wear and Cutting edge datasets.

Classification as low-high (L-H) wear

The results obtained using a binary classification are better than those obtained in a three class classification. In order to perform this experiment, the low and medium classes were merged into only one class called simply low, while the high class was kept. Figure 6.15 shows the results. The best performance was achieved using ALBP with LOSIB with 8 neighbours on the Edge Wwar dataset, as in the previous experiment with three classes. However, in this case hit rate increases up to 74.04%, which improves in 38.37% the results for three classes with the same configuration.

6.4.3. Wear region dataset results

Results with LBP methods

The first experiment consisted of evaluating our dataset using just LBP and CLBP. In that case we have omitted the ALBP results due to its bad performance.

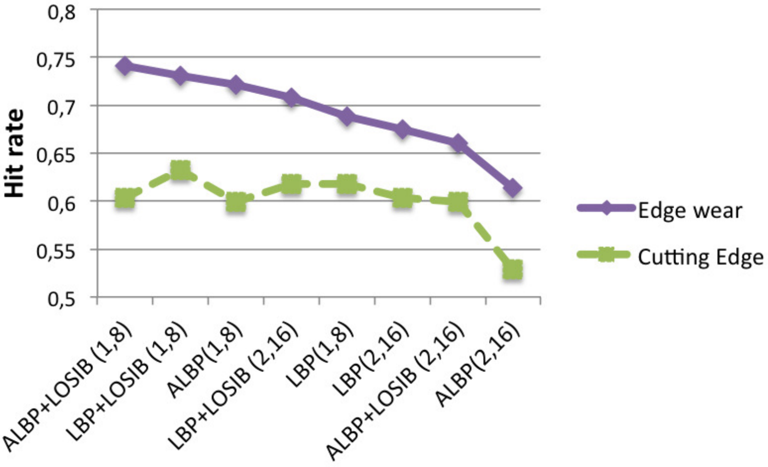


Figure 6.15: Results using low-high classification on Edge wear and Cutting edge datasets.

Furthermore, LOSIB descriptor was used to describe by itself the images just to have a briefly idea of its own performance. In Figure 6.16, we can see a graphic with the results for the L-H classification. As it can be seen, all classical methods except the CLBP with 16 neighbours obtain better results than the LOSIB method by itself. In this dataset the use of a little neighbourhood shows better results in the LBP descriptors due to the narrow shape of the region of interest. LBP with 8 neighbours and radius 1 achieved a 78.44% of hit rate in contrast with LOSIB (2,16) that had a performance of 74.57%. In Figure 6.17, the results achieved by the three classes clas-

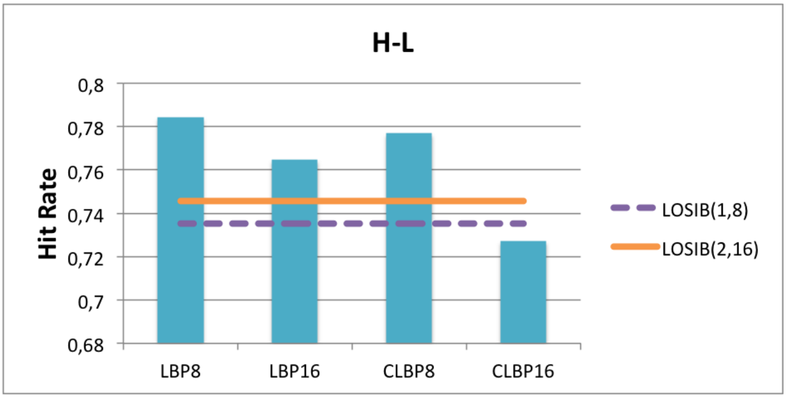


Figure 6.16: Results using SVM and classical LBP and CLBP with different neighbourhoods for the binary classification (H-L). The horizontal lines represent the results achieved with LOSIB.

sification (L-M-H) are shown. It can be observed that the performance using this classification decreases due to the difficulty to discern among three classes instead of just two. In this experiment all the methods obtained better results using the small neighbourhood in contrast with the binary classification, where LOSIB(2,16) achieved better results than LOSIB(1,8). The best result is again obtained when using LBP8, with a 67.30% of hit rate, while LOSIB(1,8) achieved just a 59.25%. However, LOSIB was developed not for describing an image by itself but for boosting the description of other methods. In the next section the results obtained when LBP methods are combined with LOSIB are shown.

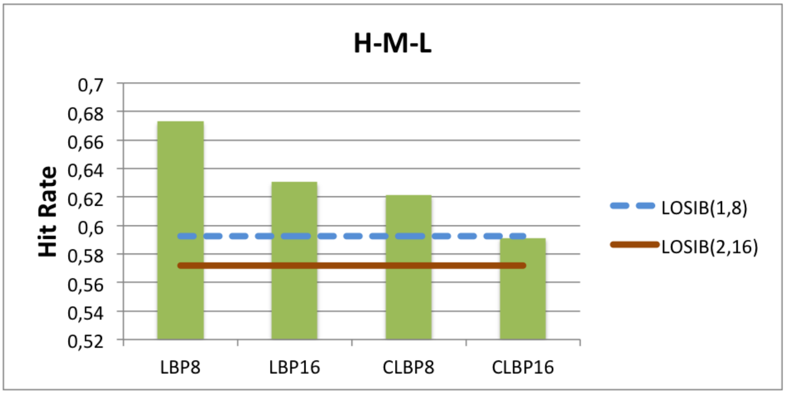


Figure 6.17: Results using SVM and classical LBP and CLBP for the ternary (H-M-L) classification. The horizontal lines represent the results achieved with LOSIB.

Results with LBP+LOSIB methods

An experiment combining LOSIB with LBP based methods was carried out in order to determine the real performance of LOSIB as a booster. In Figure 6.18, a visual comparison is depicted to better understand the results in the binary classification. The horizontal lines show the four values obtained in the previous experiment, using just LBP and CLBP with 8 and 16 neighbours. The colour of the bars is chosen to show the improvement of each fusion method with its base descriptors. It can be observed that in all cases the combination of LBP variant with LOSIB improves the performance of classification. In the case of CLBP16+LOSIB(1,8) the improvement is 32.92%, which is the highest difference obtained. The best result is achieved with LBP8+LOSIB(1,8) yielding an 80.58% of hit rate, and an improvement of 2.73%.

For the ternary classification, we obtained similar results (see Figure 6.19). Also, the best result was achieved when using LBP8+LOSIB(1,8) with a 67.76% of hit rate. The highest improvement was obtained combining CLBP8 with LOSIB(1,8) obtaining a 67.36% of hit rate, better than the baseline CLBP8 method which achieves just

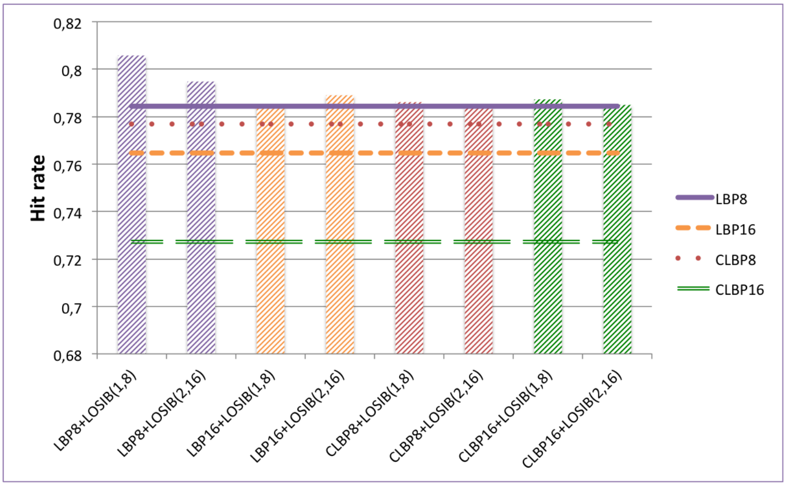


Figure 6.18: Results using SVM and fusion of LBP and CLBP with LOSIB for the binary (H-L) classification.

62.3% of hit rate. The results using three classes are worse than the results for two classes but they also give more information about the wear level of the inserts.

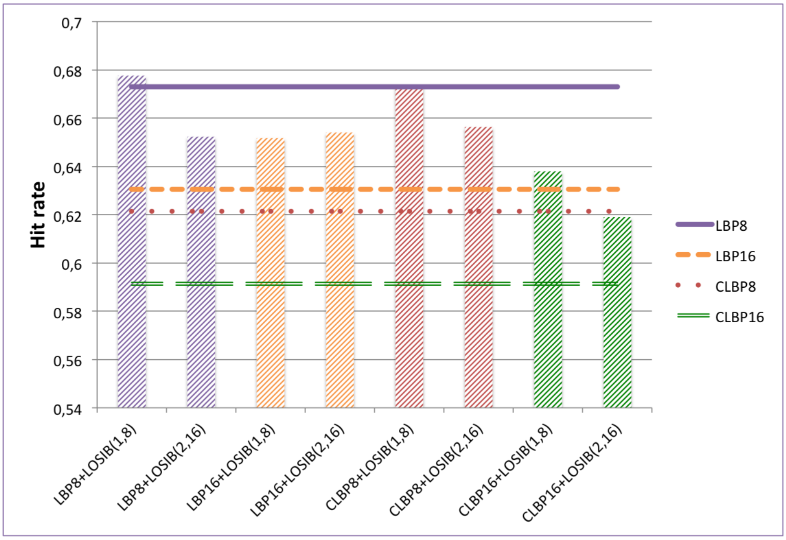


Figure 6.19: Results using SVM and fusion of LBP and CLBP with LOSIB for the ternary (H-M-L) classification.

6.5. Conclusions

A new texture booster has been developed in order to enhance global and local descriptors (e.g. Wavelet-based feature vectors and Local Binary Pattern, respectively). Local Oriented Statistical Information Booster (LOSIB) extracts the local oriented information of the image, taking into account the means of the gray value differences of the pixels and their neighbours along different orientations. Two different parameters: radius of the neighbourhood and number of neighbours, give LOSIB more reliability and robustness than ALBPS adapting the method to different kinds of images and problems. Several classical methods have been evaluated alone and also concatenated with LOSIB with three different parameter combinations: the first one, using eight neighbours with a depth of one pixel; the second one using sixteen neighbours with a depth of two pixels and the third one combining both of them. Results have shown that the early fusion of the descriptors with LOSIB increases the performance of the classification in all the different experiments we have carried out. In material recognition, two datasets have been used, KTH-TIPS 2a and Brodatz32, obtaining better results in all the cases except when combining LBP with LOSIB(1,8) and LOSIB(2,16) separately, but outperforming always the hit rate when both of them were concatenated together. Two different classification methods have been tested in order to give more credibility to the results. A one-vs-one SVM approach with a polynomic kernel of order two and a weighted kNN variant. The best results were achieved with the CLBP method combined with LOSIB(1,8) obtaining a 71.44% of hit rate in KTH-TIPS 2a and using LBPV and LOSIB(1,8)+LOSIB(2,16) on Brodatz32 with a 91.06% of hit rate. While the best performance was achieved using local descriptors, the best improvement was obtained when LOSIB was combined with global descriptors, as they complement each other mixing local and global information.

The second experiment deals with the classification of the boar acrosome integrity as intact or damaged using gray scale digital images. Three descriptors have been considered, compared and combined. We have computed 13 Haralick features on the GLCM of the original image and the four sub-bands of the first split of the Haar DWT, making up a global texture descriptor of the head of the spermatozoon. Besides, a local texture description is obtained by computing a rotation invariant and uniform LBP. Finally, the shape is described through FSD giving an idea of the contour of the spermatozoon head. We use a early fusion approach, concatenating the three selected descriptors before the classifier since each one provides new descriptive information. A k-fold cross validation analysis has been implemented to make the results independent from the classification parameters selection. The classification has been carried out with a SVM backed by a LS algorithm and four metrics have been computed in order to measure the performance of our proposal: precision, recall, F-Score and accuracy. Results have shown that the best performance is achieved by the combination of the three selected descriptors (WFLP8) with a distance of 6 pixels in the GLCM matrix and a radius 1 and 8 neighbours in the LBP. The 99.19% of accuracy and 0.9913 of F-Score makes this approach the best to this date Alegre et al. (2013) Alegre et al. (2012) Alegre et al. (2008) and very advantageous for the veterinary community. Furthermore, we have tested the combination of LOSIB with the previous descriptors confirming the good performance of our proposal. With WFLP16 and LOSIB, we achieved the best results ever obtained with a 99.73% of accuracy and a 0.9975 of FScore. Moreover, the dataset is much bigger than in previous works with a total of 1851 heads, making the results more reliable.

The third research field evaluated was tool wear estimation. We carried out two different experiments: In the first one, two types of dataset were used following two strategies. The first strategy consisted of a ternary classification with low, medium and high wear labels. The second strategy was a binary classification between low and high wear labels. In a first phase, complete edges were used to describe their texture using ALBP, LBP combined with LOSIB and using different neighborhood configuration. The best results were obtained using the binary classification with LBP and LOSIB(1,8) achieving a 63.21% of hit rate. For the ternary classification the best result was 43.40% of hit rate using ALBP with LOSIB(1,8). In a second phase, texture methods evaluation was done using the Edge Wear dataset, that is, focusing only in the worn region. Best results here were achieved in the ternary classification with a 53.77% of hit rate, which outperforms the first phase results with the complete edge in 23.89%. In the binary classification hit rate was 74.06% using ALBP with LOSIB(1,8), that is, a 22.65% of improvement with regard to the first phase.

The second experiment was developed taking into account the wear dataset which divides the insert images into one, two or three subimages depending on the different wear regions appeared in them. In the binary classification, the best results were achieved using LBP8 and LOSIB(1,8) with a 80.58% of hit rate, outperforming the based methods in more than a 2.7%. It is noticeable that in all the experiments the addition of LOSIB information improves the accuracy of the classification. When using CLBP, the highest increment was achieved combining LOSIB(1,8) with CLBP16, obtaining a 7.95% of improvement. In the ternary classification, more challenging than binary classification, the best result was obtained also using LBP8 and LOSIB(1,8) with a 67.76% of hit rate, outperforming the LOSIB combination of all the based methods as in the binary one.

Both classification experiments are very interesting for manufacturing engineering in order to implement an automatic control process. Results are very promising due to the opportunity to determine the optimal time when an insert requires to be replaced. This can lead to significant savings in time with regard to traditional methods of manual checking or machined part/cutting time counting. Also, it can prevent from accidental tool breakage and part damage.

In conclusion, a new local texture booster has been developed which combined with all the studied descriptors outperforms the classification in all cases resulting in a very promising method.

Chapter 7

Complete Local Oriented Statistical Information Booster (CLOSIB)

7.1. Methodology

We proposed a generalisation of our previos method, LOSIB, and we named it Complete Local Oriented Statistical Information Booster (CLOSIB). CLOSIB aims at improving the description performance of image feature descriptors.

CLOSIB is computed from the statistical information of the grayscale gradient magnitude of each pixel of the image. The gradient information of an image has been used in several texture descriptors in the state of the art. However, the statistical information of the gradient magnitudes is infrequently taken into account for the description of an image. CLOSIB is conceptually simple and straightforward to implement. We derive CLOSIB for a general case based on a circularly symmetric neighbor set of P members on a circle of radius R. Parameter R determines the spatial resolution of the booster, whereas P controls the quantization of the angular space. Equation (7.1) defines the CLOSIB of an image.

$$CLOSIB_{P,R,\theta} = \mathop{\Big|\Big|}_{p=1}^{P/\eta} \left((\theta - 1)\mu_2^p - (-1)^\theta (\mu_1^p)^\theta\right)^{1/\theta} \tag{7.1}$$

where $||$ represents the concatenation function, $\theta \in \{1, 2\}$ is the order of the statistical moment considered, μ_1^p and μ_2^p are, respectively, the first and second statistical raw moments determined by Equation (7.2) and η is a factor that controls the portion of the considered orientations in the quantized angular space. If not specified, we set $\eta = 1$. Therefore, CLOSIB is a feature vector of dimensionality P/η.

$$\mu_i^p = \frac{1}{N} \sum_{c=1}^{N} \left(d_p(x_c, y_c)\right)^i \tag{7.2}$$

where N is the number of pixels in the image and $d_p(x_c, y_c)$ is the absolute difference of the gray value of the center pixel g_c, located in (x_c, y_c), and the gray value of the neighbor g_p, located at a distance R with orientation $2\pi p/P$ from the center pixel, as stated in Equation (7.3).

$$d_p(x_c, y_c) = |g_c - g_p| \tag{7.3}$$

The coordinates (x_p, y_p) of a neighbor located at a distance R with orientation $2\pi p/P$ from the center pixel are obtained by means of Equation (7.4).

$$(x_p, y_p) = (x_c + R\cos(2\pi p/P),\ y_c - R\sin(2\pi p/P)) \tag{7.4}$$

The gray value of neighbors that are not located in the centres of pixels is estimated by interpolation of their connected pixels.

Figure 7.1 shows an example of the absolute differences of the gray values $d_p(x_c, y_c)$ for $P = 8$ orientations in a neighborhood of radii $R = 1$ and $R = 2$.

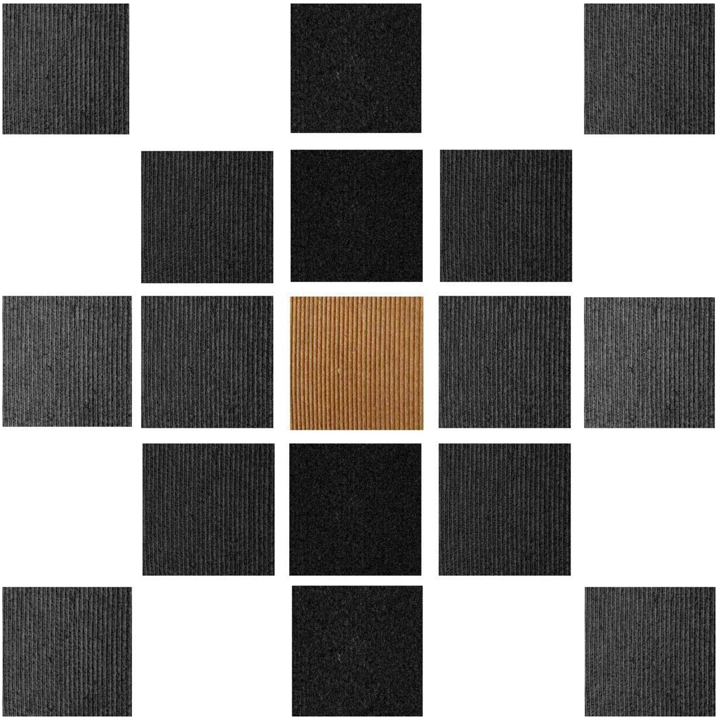


Figure 7.1: Absolute differences of the gray values $d_p(x_c, y_c)$ for $P = 8$ orientations in a neighborhood of radii $R = 1$ and $R = 2$. The original image is shown in the center. The main change in the intensity of the original image occurs in the horizontal direction $p = 1$ and $p = 5$.

CLOSIB allows to adjust three parameters: the order of the statistical moment θ, the radius of the neighborhood R and the quantization of the angular space P.

The order of the statistical moment, θ, determines the statistical measure that is used to compute CLOSIB. When $\theta = 1$, CLOSIB is a feature vector whose elements are the means of the absolute differences of the gray values $d_p(x_c, y_c)$ for each orientation and every pixel in the image. In the case of $\theta = 2$, the elements of CLOSIB are the standard deviations of the absolute differences of the gray values $d_p(x_c, y_c)$. In relation to the radius of the neighborhood R, small radii are quite effective in images with a high level of heterogeneity. As the size of the neighborhood increases, noise is reduced but at the expense of a possible loss of valuable information, specially in images with high variability of the pixel values.

Regarding the quantization of the angular space P, a higher value of P means that a greater number of orientations are considered in the computation of CLOSIB. As the texture becomes more heterogeneous, the number of orientations should increase in order to capture all the variety of the image. However, using an excessive number of orientations on homogeneous textures may be counter-productive due to the loss of weight of the important ones.

7.1.1. CLOSIB variants

In the literature, LBP is typically computed for (P, R) pairs of values equals $(8, 1)$, $(16, 2)$ or a concatenation of both. Likewise, CLOSIB can be computed for (P, R, θ) triples of values equals $(8, 1, 1)$, $(8, 1, 2)$, $(16, 2, 1)$, $(16, 2, 2)$ or a concatenation of several of them. CLOSIB can also be computed for any other triple of values. We indicate the concatenation of several CLOSIBs with the symbol ||. For example, the concatenation of CLOSIB$_{8,1,1}$ and CLOSIB$_{16,2,1}$ is represented as CLOSIB$_{8,1,1\,||\,16,2,1}$.

In this section, we propose and describe three specific ways of obtaining CLOSIB.

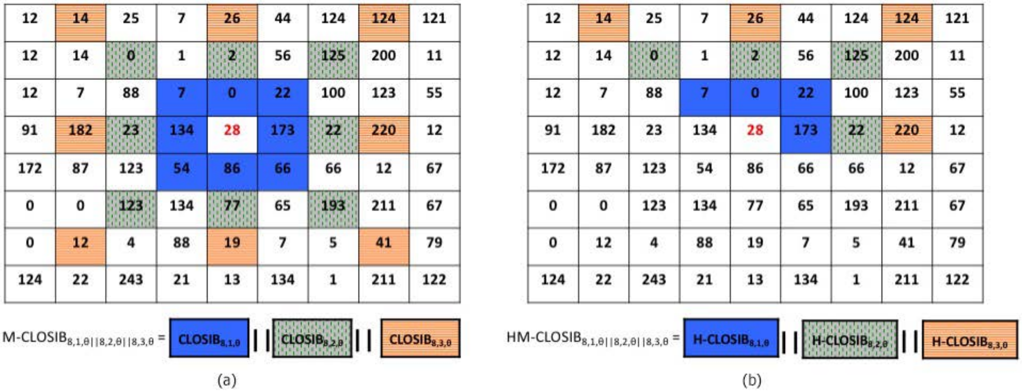
Multi-scale CLOSIB (M-CLOSIB)

In Chan et al. (2007) proposed a multi-scale LBP (MSLBP) method for face detection that benefits from the multi-resolution information captured from the regional histogram. MSLBP has been extended and applied to other fields in the literature (Jia et al., 2014; Wen et al., 2016).

Similarly, we introduced a multi-scale CLOSIB which we named M-CLOSIB. M-CLOSIB is a concatenation of the CLOSIBs obtained for a fixed number of orientations P and several radii of the neighboorhoud R. Figure 7.2a shows an schema of the computation of M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$.

Half CLOSIB (H-CLOSIB)

For even values of P, CLOSIB encompasses statistical information of the absolute differences of the gray values $d_p(x_c, y_c)$ along directions that differ in π radians. Figure 7.3a shows this fact for $P = 8$. The statistical information along directions that differ in π radians is usually very similar. Figure 7.3(b-c) illustrates two examples.



12	14	25	7	26	44	124	124	121
12	14	0	1	2	56	125	200	11
12	7	88	7	0	22	100	123	55
91	182	23	134	28	173	22	220	12
172	87	123	54	86	66	66	12	67
0	0	123	134	77	65	193	211	67
0	12	4	88	19	7	5	41	79
124	22	243	21	13	134	1	211	122

M-CLOSIB$_{8,1,\theta||8,2,\theta||8,3,\theta}$ = CLOSIB$_{8,1,\theta}$ || CLOSIB$_{8,2,\theta}$ || CLOSIB$_{8,3,\theta}$

(a)

12	14	25	7	26	44	124	124	121
12	14	0	1	2	56	125	200	11
12	7	88	7	0	22	100	123	55
91	182	23	134	28	173	22	220	12
172	87	123	54	86	66	66	12	67
0	0	123	134	77	65	193	211	67
0	12	4	88	19	7	5	41	79
124	22	243	21	13	134	1	211	122

HM-CLOSIB$_{8,1,\theta||8,2,\theta||8,3,\theta}$ = H-CLOSIB$_{8,1,\theta}$ || H-CLOSIB$_{8,2,\theta}$ || H-CLOSIB$_{8,3,\theta}$

(b)

Figure 7.2: Schema that shows the neighborhood around a given pixel, marked with red font, considered for the computation of M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ (a) and HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ (b). The pixels marked in blue represent the neighborhood for M-CLOSIB$_{8,1,\theta}$ (a) and HM-CLOSIB$_{8,1,\theta}$ (b). The pixels marked in gray represent the neighborhood for M-CLOSIB$_{8,2,\theta}$ (a) and HM-CLOSIB$_{8,2,\theta}$ (b). The pixels marked in pink represent the neighborhood for M-CLOSIB$_{8,3,\theta}$ (a) and HM-CLOSIB$_{8,3,\theta}$ (b). M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ is obtained as a concatenation of CLOSIB$_{8,1,\theta}$, CLOSIB$_{8,2,\theta}$ and CLOSIB$_{8,3,\theta}$. HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ is obtained as a concatenation of H-CLOSIB$_{8,1,\theta}$, H-CLOSIB$_{8,2,\theta}$ and H-CLOSIB$_{8,3,\theta}$.

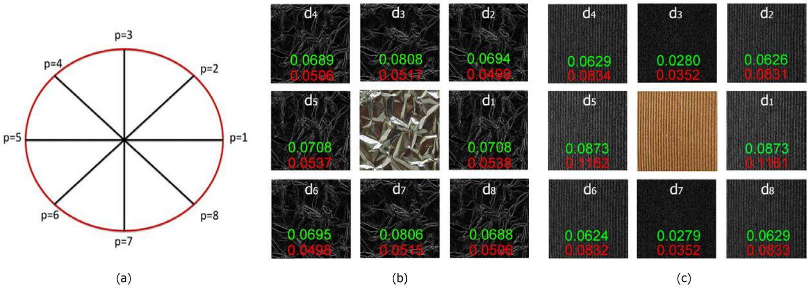


Figure 7.3: (a) Circumference that represents the neighborhood considered for the computation of CLOSIB with $P = 8$.(b) and (c) Schemas that represent the computation of CLOSIB$_{8,1,\theta}$ for two different images.

We define a Half CLOSIB (H-CLOSIB) following Equation 7.1 with $\eta = 2$. The angular space is yet quantized in P equal parts but only the first P/η orientations are taken into account for the computation of H-CLOSIB. Figure 7.4 shows an example of the orientations considered when computing CLOSIB$_{8,1,\theta}$ and H-CLOSIB$_{8,1,\theta}$.

H-CLOSIB presents two main characteristics. First, it may eliminate redundant statistical information. Second, the dimensionality of H-CLOSIB is half of the dimensionality of the equivalent CLOSIB. This might be decisive in some cases when

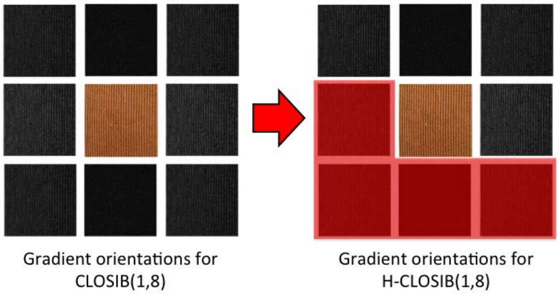


Figure 7.4: Schemas of the computation of CLOSIB$_{8,1,\theta}$ (a) and H-CLOSIB$_{8,1,\theta}$ using the example of Figure 7.3(c).

the amount of memory or computational time are critical, such as in embedded systems with little RAM.

Half Multi-scale CLOSIB (HM-CLOSIB)

We propose a Half Multi-scale CLOSIB (HM-CLOSIB) which is obtained as a M-CLOSIB when $\eta = 2$. This variant combines the advantages and disadvantages of both M-CLOSIB and H-CLOSIB. Figure 7.2b shows an schema of the computation of HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$.

In tables 7.1 and 7.2, we show the CLOSIBs that we used in the experiments for CLOSIB and H-CLOSIB, and M-CLOSIB and HM-CLOSIB, respectively.

Table 7.1: Each row describes the parameters used to compute different CLOSIBs and H-CLOSIBs in the experiments. Both values, 1 and 2, in the column 'order' indicates that we obtained CLOSIB as a concatenation of the CLOSIBs for each statistical moment, $CLOSIB_{P,R,1\,||\,P,R,2}$.

Radius (R)	Neighbors(orientations) (P)	Order (θ)
1	8	1
1	8	2
2	16	1
2	16	2
1	8	1,2
2	16	1,2

Table 7.2: Each row describes the parameters used to compute different M-CLOSIBs and HM-CLOSIBs in the experiments. Several values for a parameter indicate that we obtained CLOSIB as a concatenation of the CLOSIBs for each single value.

Radius (R)	**Neighbors(orientations) (P)**	**Order (θ)**
1,2,3	8	1
1,2,3,4,5	8	1
1,2,3	8	2
1,2,3,4,5	8	2
2,3,4	16	1
2,3,4,5,6	16	1
2,3,4	16	2
2,3,4,5,6	16	2
1,2,3	8	1,2
1,2,3,4,5	8	1,2
2,3,4	16	1,2
2,3,4,5,6	16	1,2

7.2. Material Recognition

7.2.1. Experimental Setup

For KTH-TIPS 2a dataset, we used the experimental protocol developed by Caputo et al. (2005); Chen, Tian, Lee, Zheng, Smith and Laine (2010). We used all images of one sample of each material for testing and the rest for training. This experimental setup is more challenging than a random division of the images into training and test sets due to the high inter-sample variation. We repeated this experiment four times with different samples for testing. Since there are four samples for each material in the dataset, we used a different sample for testing in each experiment. We report the results as the average hit rate over the four runs. We define hit rate as the number of correctly classified images divided by the total number of images in the test set. We used a Support Vector Machine (SVM) with the one-vs-one paradigm (Hsu and Lin, 2002) to classify the images with the Least Squares training algorithm and a polynomial kernel of order 2.

7.2.2. Results

CLOSIB versus LBP-based descriptors

We developed CLOSIB as an enhancer of texture descriptors. However, in this section we show the performance of CLOSIB as a descriptor itself. Figure 7.5 presents the results that we obtained when describing the images with CLOSIB and with descriptors based on LBP. For all CLOSIB variants, we achieved the best performance using a concatenation of the CLOSIBs for the first and second statistical moments. For all LBP-based descriptors, we obtained the best results for $R = 2$ pixels and $P = 16$ neighbors. It is remarkable that we achieved the highest performance using HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ which yielded a hit rate of 67.93%. Therefore, the proposed enhancer by itself outperforms some of the state of the art LBP-based descriptors.

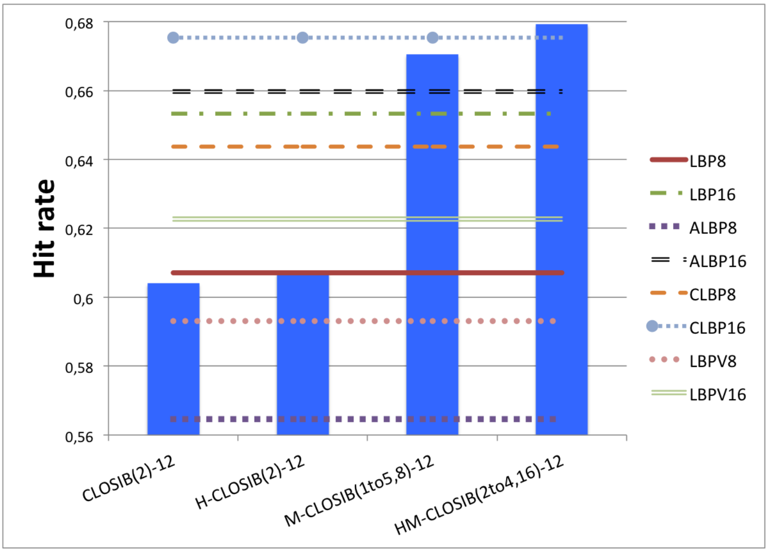


Figure 7.5: Hit rates when we describe KTH Tips2-a images with different CLOSIBs and LBP-based descriptors. For each CLOSIB variant -CLOSIB (standard), M-CLOSIB, H-CLOSIB and MH-CLOSIB-, we only represent the best result obtained among the results with different combinations of parameters.

CLOSIB + LBP-based descriptors

The following experiment consists of combining LBP-based descriptors with CLOSIB. The combination is done by means of a concatenation. Figure 7.6 and Table 7.3 graphically and numerically show the results.

In all experiments, we achieved the best results using CLOSIB as an enhacer of LBP-based descriptors in opposition to only using LBP-based descriptors. We obtained the highest hit rates equals to 72.54% with CLBP$_{16,2}$ ||HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ and CLBP$_{16,2}$ ||H-CLOSIB$_{16,2,1\,||\,16,2,2}$ closely followed by LBP$_{16,2}$ ||MH-CLOSIB$_{8,1,1\,||\,8,1,2\,||\,8,2,1\,||\,8,2,2\,||\,8,3,1\,||\,8,3,2\,||\,8,4,1\,||\,8,4,2\,||\,8,5,1\,||\,8,5,2}$ with a hit rate of 72.50%. For 6 out of the 8 LBP-based descriptors, we achieved the best results with the concatenation of HM-CLOSIB.

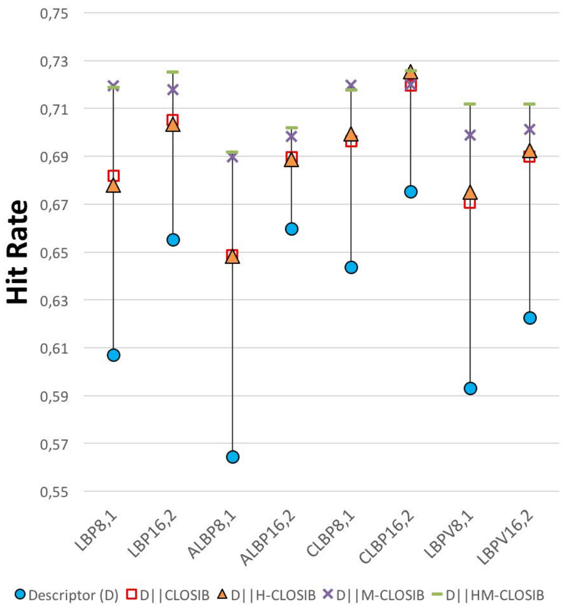


Figure 7.6: Hit rates obtained with a given LBP-based descriptor (LBP, ALBP, LVPV and CLBP) and the concatenations of the descriptor with CLOSIB variants.

Table 7.3: Hit rates (in %) obtained with a given LBP-based descriptor (LBP, ALBP, LVPV and CLBP) and the concatenations of the descriptor with CLOSIB variants. The best results for each LBP-based descriptor are highlighted in bold. The best overall results are underlined.

Descriptor (D)	**D**	**D || CLOSIB**	**D || H-CLOSIB**	**D || M-CLOSIB**	**D || HM-CLOSIB**
LBP$_{8,1}$	60.71	68.20	67.80	**71.95**	71.86
LBP$_{16,2}$	65.53	70.52	70.33	71.78	**72.50**
ALBP$_{8,1}$	56.46	64.86	64.84	68.97	**69.15**
ALBP$_{16,2}$	65.97	68.96	68.88	69.84	**70.16**
LBPV$_{8,1}$	59.30	67.05	67.51	69.89	**71.15**
LBPV$_{16,2}$	62.27	69.00	69.24	70.14	**71.17**
CLBP$_{8,1}$	64.37	69.63	69.95	**71.97**	71.76
CLBP$_{16,2}$	67.53	71.95	**<u>72.54</u>**	72.01	**<u>72.54</u>**

No multi-scale versus multi-scale LBP-based descriptors

The good performance of multi-scale CLOSIB lead us to reproduce the experiments for multi-scale LBP-based descriptors. Figure 7.7 shows the comparison between the results obtained with LBP-based descriptors and their multi-scale versions. We defined a multi-scale LBP as a concatenation of the LBP descriptors obtained with different neigborhood radii (R) and the same number of neighbors (P). The hit rate fairly improves with multi-scale LBP-based descriptors in all cases.

Therefore, multi-scale descriptors are very interesting for texture retrieval. Best result was obtained with CLOSIB$_{16,2\,||\,16,3\,||\,16,4}$ with a 71.28%. This result means a 5.55% of improvement compared with the standard CLOSIB$_{16,2}$. However, our proposed descriptor HM-CLOSIB combined with CLBP$_{16,2}$ gets the best performance until now.

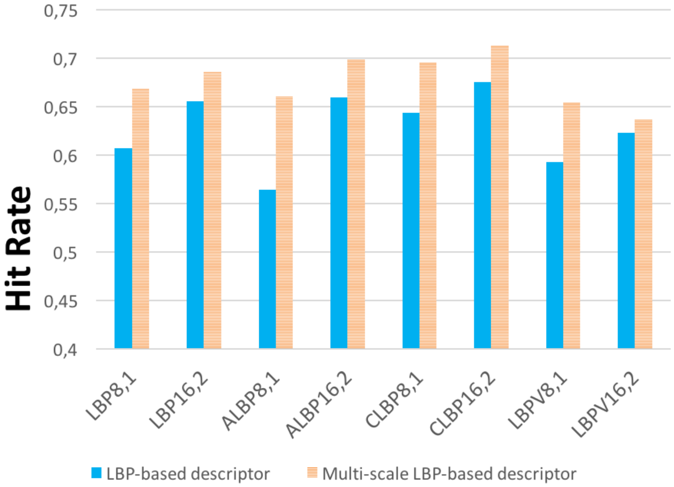


Figure 7.7: Hit rates for the LBP-based descriptors (LBP, ALBP, CLBP and LBPV) and their multi-scale versions.

HM-CLOSIB + Multi-scale LBP-based descriptors

Finally, we evaluated the combination of multi-scale LBP-based descriptors with HM-CLOSIB. We selected HM-CLOSIB due to the high performance achieved in terms of accuracy and computational time in previous experiments.

Figure 7.8 shows the hit rate of the concatenation of multi-scale LBP-based descriptors with HM-CLOSIB. Furthermore, we also present the hit rate of the (non multi-scale) LBP-based descriptors combined with HM-CLOSIB to represent the improvement in accuracy. CLBP$_{16,2}$ ||HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ outperformed the rest of the methods with a hit rate of 74.83%, which represents an improvement of at least 3.16% in hit rate with respect to the rest of descriptors.

Comparative with the state-of-the-art

Several authors tested their algorithms using KTH TIPS2-a dataset. In Table 7.4, we can see the results achieved in several papers dating from 2010. In all cases, the proposed descriptor CLBP$_{16,2}$ ||HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$

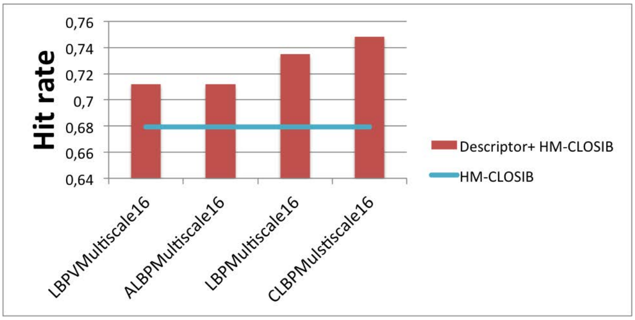


Figure 7.8: Hit rates obtained with the concatenation of multi-scale LBP-based descriptors and HM-CLOSIB. The horizontal line represents the hit rate of HM-CLOSIB descriptor.

Table 7.4: Hit rates obtained in state-of-the-art papers using KTH TIPS2-a.

Descriptor	Hit Rate (%)	Year	Reference
WLD	56.4	2010	(Chen, Tian, Lee, Zheng, Smith and Laine, 2010)
MWLD	64.7	2010	(Chen, Tian, Lee, Zheng, Smith and Laine, 2010)
LHS	73.0	2012	(Sharma et al., 2012)
CMLBP	73.1	2012	(Li and Fritz, 2012)
LQP	64.2	2012	(Hussain and Triggs, 2012)
CMR	69.4	2013	(Zhang et al., 2013)
PC	71.5	2013	(Zhang et al., 2013)
WLBP	64.4	2013	(Liu et al., 2013)
DRLTP	62.6	2014	(Satpathy et al., 2014)
DRLBP	59.0	2014	(Satpathy et al., 2014)
AMBP	70.3	2015	(Hafiane et al., 2015)
CRDP$_{3D}-2$ **(NNC)**	73.8	2017	(Wang et al., 2017)
HM-CLOSIB	67.9		
CLBP$_{16,2}$ **|| HM-CLOSIB**	**74.8**		

outperforms previous works. Furthermore, using just HM-CLOSIB as descriptor, we yielded higher hit rate than several of these studies.

7.3. Face Recognition

7.3.1. Experimental Setup

Regarding JAFFE dataset, we used the same evaluation setup proposed by Sharma et al. (2012). Specifically, one random image of each facial expression and person forms the test set and the rest define the training set. We repeat the classification 10 times to avoid biased results due to the random process. We used the multi-block approach introduced by Zhang et al. (2007) for describing a face using LBP-based descriptors and CLOSIB. We split the image into 8×8 blocks and compute a descriptor for each block. We define the descriptor of the image as the concatenation of the descriptors of the blocks. We performed two sets of experiments with JAFFE dataset. On the one hand, we used the images provided in the dataset. On the other hand, we automatically cropped the face of the images using Viola-Jones method (Viola and Jones, 2001) and used the cropped images to carry out the experiments.

7.3.2. Results

We carried out two sets of experiments with JAFFE dataset: with the original images and with automatically cropped images.

Figure 7.9 shows the hit rates that we obtained in the first experiment, using the original images. In all cases, the LBP-based descriptors achieved worse results than the combination of the LBP-based descriptors with any CLOSIB variant. The combination with M-CLOSIB yielded the highest hit rates in most of the cases, except for $LBP_{16,2}$ and $LBPV_{8,1}$ in which the combination with CLOSIB outperformed the others. We achieved the best results using $LBPV_{16,2} || M\text{-}CLOSIB_{8,1,1||8,1,2||8,2,1||8,2,2||8,3,1||8,3,2||8,4,1||8,4,2||8,5,1||8,5,2}$ with a hit rate of 82.71%.

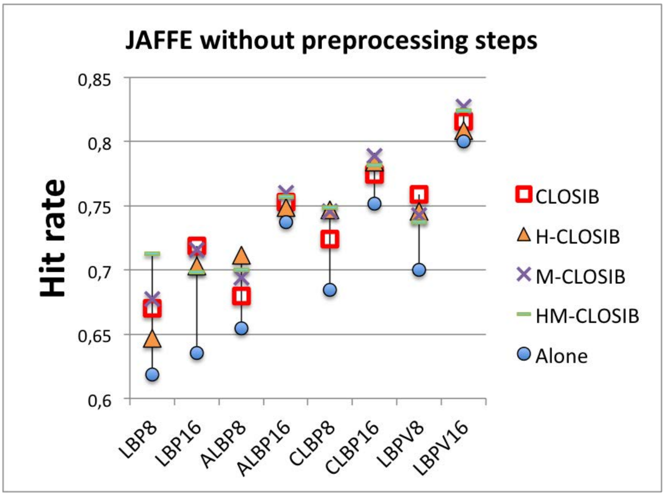


Figure 7.9: Results using the concatenation of LBP-based descriptors with CLOSIB variants (CLOSIB, H-CLOSIB, M-CLOSIB and HM-CLOSIB) on the original images of JAFFE dataset.

Regarding the second experiment, Figure 7.10 shows the hit rates achieved using the cropped images. Again, tests using LBP-based descriptors yielded worse results than when combined with any CLOSIB variant. In this case, we obtained the highest hit rate, 90.00%, using $CLBP_{16,2} || H\text{-}CLOSIB_{16,2,1||16,2,2}$. It is important to notice that carrying out the preprocessing step, the performance improves in up to 8.81%.

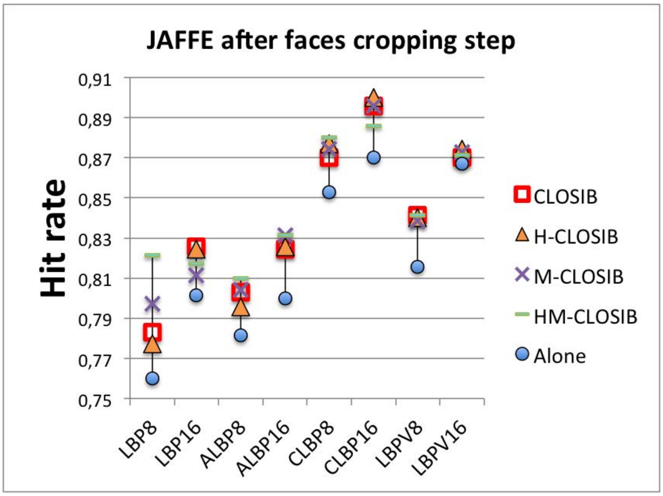


Figure 7.10: Results using the concatenation of LBP-based descriptors with CLOSIB variants (CLOSIB, H-CLOSIB, M-CLOSIB and HM-CLOSIB) on the cropped images of JAFFE dataset.

7.4. Textile recognition

7.4.1. Methodology

In Fig. 7.11 we illustrate the pipeline of our method for content-based textile retrieval. We can divide the method in two main stages: feature extraction and matching.

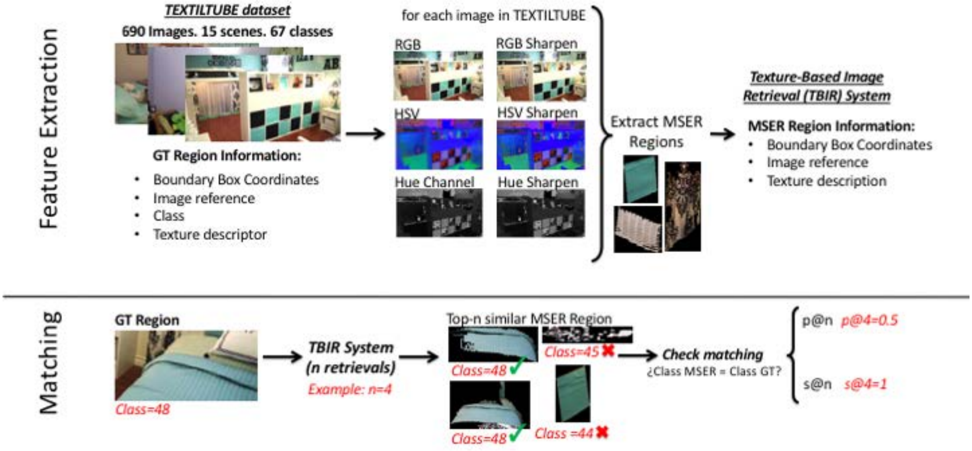


Figure 7.11: Scheme of the TBIR system.

The feature extraction is comprised of four steps. First, we convert the images to RGB and HSV colour spaces and we also extract the Hue channel. We use these three image representations for the extraction and description of the regions of interest in the images. Then, we sharp the image representations to increase the contrast along the edges where different colors meet. We adopt the unsharp masking method in which an image is sharpened by subtracting a blurred (unsharp) version of the image from itself. We use a Gaussian lowpass filter of standard deviation 1.5 for blurring the image. The maximally stable extremal regions (MSER) Matas et al. (2004) of the sharpened image representations define the regions of interest of the images. Finally, we describe the regions of interest by computing texture descriptors on the grayscale patch. We create a database in which we store for each detected region: the image coordinates of the bounding box of the region of interest in the image of reference, the images of reference themselves and the descriptors.

The matching stage allows to retrieve a given number of images that present the most similar regions to a query region (textile) of interest. It is made up of three steps. First, we describe the grayscale query region by means of the same texture descriptors. Second, we compute some distance measures among the descriptors of the query region and the descriptors of the database. Finally, the hit list is ranked by sorting the regions of the database in ascending order in relation with the distance measure.

Below, we briefly describe the methods used to build the proposed pipeline.

Region extraction: MSER

We use MSER method Matas et al. (2004) to automatically extract the regions (textiles) of interest due to the good results achieved in preliminary tests.

MSER is a method for blob detection that extracts from an image a number of co-variant regions called MSERs. These high contrast regions are connected areas characterized by almost uniform intensity, surrounded by contrasting background. MSERs are constructed by binarizing the image at multiple threshold levels and selecting the connected components that maintain their sizes over a large set of thresholds. The algorithm of MSER can be described in 4 steps:

1. For each threshold, compute the connected binary regions, i.e. extremal regions.
2. Compute the area variation for each extremal region:

$$var_i = \frac{Area_{i+\Delta} - Area_i}{Area_i} \tag{7.5}$$

   Here i represents the extremal region with maximal intensity i, and $i+\Delta$ refers to the extremal region expanded from the ith one, with maximal intensity $i+\Delta$.

var_i is the relative difference of area when maximal intensity rises from i to $i+\Delta$.

3. Find the extremal regions which present an area variation smaller than its immediate parent and ancestor in the nest, i.e. maximally stable extremal regions.

Region description

We use the following methods to describe the textiles: LBP Ojala et al. (2002), ALBP Guo, Zhang, Zhang and Zhang (2010) and HOG Dalal and Triggs (2005), and early fusion concatenations of the previous descriptors with CLOSIB and HCLOSIB enhancers.

Distance measures

We use five distance measures to compute the distances among the descriptors of the query region and the descriptors of the automatically detected regions of interest of the database. These are: Spearman, Cosine, Cityblock, Euclidean and Correlation distances.

Spearman rank correlation coefficient is a nonparametric measure of rank correlation and it measures the strength and direction of association between two ranked variables. This measure uses a variable's rank which is the average of their positions in the ascending order of the values.

Cosine distance calculates the angular cosine between two vectors following eq. (7.6).

$$d_{cos}(A,B) = 1 - \frac{AB'}{\sqrt{(AA')(BB')}} \tag{7.6}$$

Cityblock distance is calculated using eq. (7.7) and is defined by the sum of the absolute distances of every coordinate between two vectors. This measure distance depends on the rotation of the coordinate system but is invariant to reflection and translation.

$$d_{cit}(A,B) = \sum_{j=1}^{n} |A_j - B_j| \tag{7.7}$$

Euclidean distance (see eq. (7.8)) is the most commonly used distance measure and calculates the length of the straight segment which connects two vectors.

$$d_{euc}(A,B) = \sqrt{((A-B)(A-B)'} \tag{7.8}$$

The Correlation distance is obtaining by dividing the distance covariance of two vectors by the product of their distance standard deviations. See eq. (7.9)

$$d_{cor}(A,B) = 1 - \frac{(A-\overline{A}) \times (B-\overline{B})'}{\sqrt{(A-\overline{A}) \times (A-\overline{A})'}\sqrt{(B-\overline{B}) \times (B-\overline{B})'}} \tag{7.9}$$

7.4.2. Performance evaluation metrics

In retrieval systems, it is important that the retrieved images are ranked according to their relevance to the query region forming a hit list, rather than being returned as a set. The most relevant hits must be within the top images of the hit list returned for a query region. To account for the quality of ranking the hits in the hit list, we used relevance ranking measures, i.e. precision at n and success at n.

Precision at n

Precision at n, p@n, is the rate of the top-n images of the hit list correctly classified in relation to the class of the query region. Likewise, the precision at a cut-off of n elements of the hit list. We define $HitList_n$ as the set that contains the n images with smallest distance to the query region, q. Equation (7.10) presents the mathematical definition of precision at n.

$$p@n = \frac{\#H(q)}{n} \tag{7.10}$$

where $\#H(q)$ is the cardinal of $HitList_n$ in which the query class is actually present in the image. It is formally defined in equation (7.11).

$$H(q) = \{h_i/(h_i \in HitList_n) \wedge (class(h_i) = class(q)) \mid i = 1...n\} \tag{7.11}$$

where h_i is i-th retrieved image in the hit list.

Success at n

There are occasions in which the user does not need to see many relevant images but is disappointed by a completely irrelevant top-n Craswell (2009). This is the case of ASASEC project, in which finding at least one hit in all the hit list would be a satisfactory result. Success at n, s@n, measures if a relevant image was retrieved within the top-n hits of the hit list. Success at n is equal to 1 if the top-n images contain a relevant document and 0 otherwise, see equation (7.12).

$$s@n = \begin{cases} 1 & \text{if } H(q) \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{7.12}$$

where $H(q)$ is the set of images defined in equation (7.11).

7.4.3. Experimental setup

We applied the method described in Section 7.4.1 to the 690 images of TextilTube dataset, extracting a total of 58031 regions. In order to evaluate the performance of our method, we used the ground truth textile regions as query regions of interest. For each query region, we calculated p@n and s@n metrics for the retrieved hit list when computing a given distance measure among the texture descriptors of the query region and the analogous texture descriptors of the database.

7.4.4. Distance measure evaluation

In order to determine the best distance measure and present uniform results, we carried out the following voting system. For each texture descriptor in Section 7.4.1.1, we computed s@n for $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$ with all distance measures described in Section 7.4.1.3. We assigned three, two and one points to the distance measures that achieved the highest, second highest and third highest s@n, respectively, for each experiment. Finally, we summed up the points along all combinations. Figure 7.12 shows a scheme of the procedure. We disregarded a voting system that only relies on the best distance measure of each experiment because the results for the different distance measures were not enough distinctive.

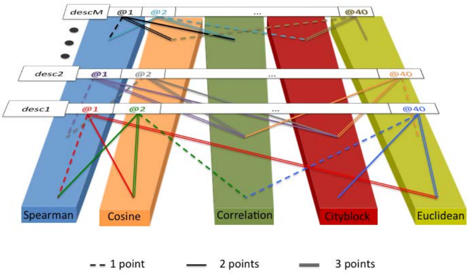


Figure 7.12: Scheme of the voting procedure to determine the best distance measure.

Figure 7.13 presents the results in parts per unity achieved with each distance measure. Correlation distance achieved the best results with a 32% of votes, followed by Cosine distance (27%) and Spearman rank correlation coefficient (20%). The commonly used Euclidean distance only yielded a 7% of the votes. Therefore, we carried out our experiments using the Correlation distance.

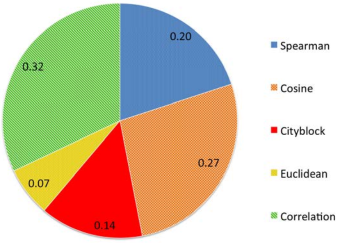


Figure 7.13: Results of the voting process in parts per unity for the different distance measures.

7.4.5. Results

In this section we present the results obtained following the proposed method and experimentation for each evaluated texture descriptor.

Figure 7.14 shows the precision at n (p@n) achieved for all texture descriptors. We used values of $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$. For $n \leq 19$, HOG+HCLOSIB outperformed the rest with a precision of 37.17% for $n = 1$. The early fusion of CLOSIB and HCLOSIB with HOG outperforms HOG alone. However, the early fusion of CLOSIB and HCLOSIB with LBP obtained the worst results. In the case of ALBP, the descriptor alone outperforms the early fusion for small values of n whereas the opposite is true for high values of n. It is notorious the better performance of ALBP (28.83% for $n = 1$) versus LBP (16.60% for $n = 1$).

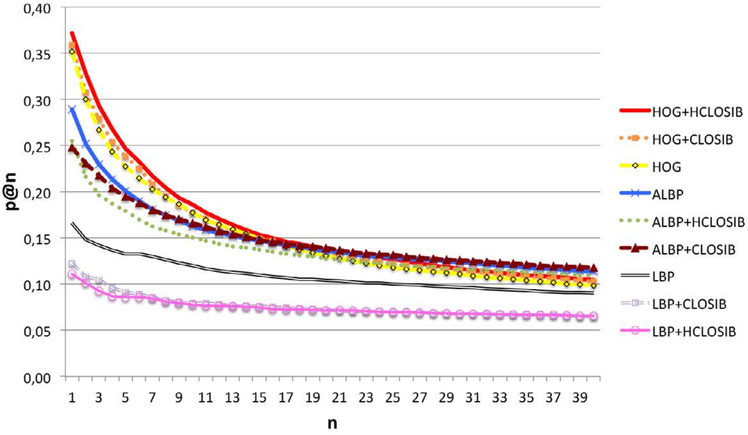


Figure 7.14: Precision at n (p@n) for all texture descriptors using Correlation distance and $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$.

Figure 7.15 illustrates the success at n (s@n). For values of $n \leq 8$, HOG+HCLOSIB yielded the best results, whereas for higher values of n, ALBP outperformed the others with a 82.00% of s@40 (74.86% with HOG+HCLOSIB). LBP achieved low s@n at small values of n but it is the second best descriptor for $n = 40$ reaching 76.04% of success. CLOSIB enhancer improves the performance of HOG and decreases the performance of LBP and ALBP.

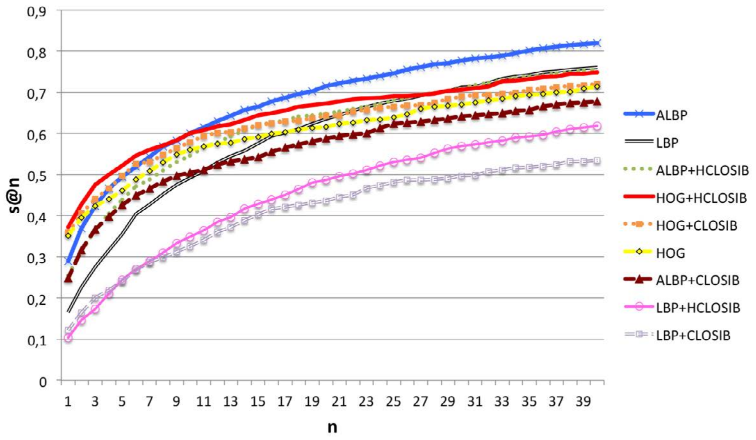


Figure 7.15: Success at n (s@n) for all texture descriptors using Correlation distance and $n \in \mathbb{N} \mid n = \{1, 2, \ldots 40\}$.

In order to get a unique value to evaluate the performance of each descriptor, we computed the arithmetic mean of the success and precision for three intervals of $n \in \mathbb{N} \mid n = \{1, 2, \ldots j\}$, where $j = \{10, 20, 40\}$. Figure 7.16 shows the arithmetic mean of the success and the precision in these intervals of values of n. Regarding p@n, HOG+HCLOSIB outperformed the rest of descriptors in all the intervals. With respect to s@n, HOG+HCLOSIB yielded the best results for the intervals with $j = 10$ and $j = 20$ whereas ALBP did for $j = 40$.

7.5. Conclusions

We proposed a new texture descriptor booster, called CLOSIB, which is based on the statistical information provided by the image gradient. Furthermore, we presented three variants of CLOSIB: H-CLOSIB, useful for embedded systems or machines with a low RAM memory; M-CLOSIB, a multi-scale descriptor which extracts information for consecutive neighborhoods; and HM-CLOSIB, which is a multi-scale H-CLOSIB. We evaluated CLOSIB in two applications: material recognition using

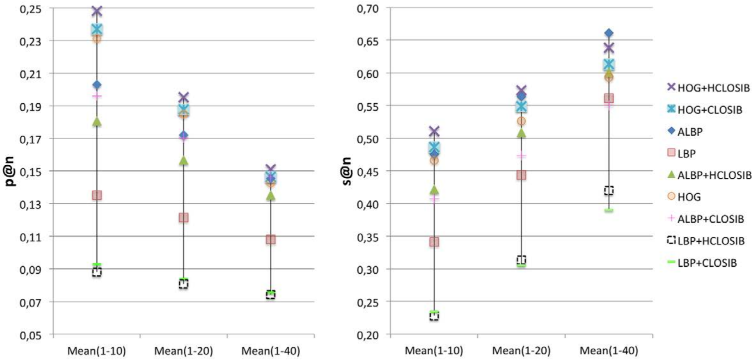


Figure 7.16: Arithmetic mean of precision and success at n for intervals of n from 1 to 10, from 1 to 20 and from 1 to 40.

KTH TIPS2-a dataset and face recognition using JAFFE dataset.

Regarding material recognition, HM-CLOSIB outperformed some of the state-of-the-art LBP-based descriptors. To check the performance of CLOSIB as an enhancer of other texture descriptors, we used a concatenation of LBP-based descriptors with CLOSIB variants. All tested combinations of LBP-based descriptors with CLOSIB yielded better results than the individual descriptors. Moreover, we proved that the classification results for material recognition improves when using multi-scale LBP-based descriptors. We obtained the best result using a concatenation of a multi-scale CLBP and HM-CLOSIB yielding a hit rate of 74.83%. Finally, this method outperformed relevant state-of-the-art methods tested on KTH TIPS2-a images.

In relation to face recognition, the combination of LBP-based descriptors with CLOSIB variants outperformed the individual descriptors as well. We obtained the highest hit rate of 90% using a combination of $\text{CLBP}_{16,2}$ and H-CLOSIB when automatically cropping the images of the dataset by means of Viola-Jones method.

Furthermore, we have proposed a novel method for textile retrieval based on texture description of the most stable regions of the images.

For this application, we have used MSER method to detect all the possible regions connected with a specific textile and after that, different methods based on texture such as LBP, ALBP or HOG and CLOSIB used in combination with the other three methods for the textile description used in the retrieval process.

Several experiments using different distances have demonstrated than the most discriminant one was the correlation distance. Furthermore, the better description methods have been ALBP and HOG+HCLOSIB using success at n and precision at n respectively. Taking into account just the most similar image retrieved, HOG+HCLOSIB achieved a precision of 37.17% which is very interesting due to the number of classes in the dataset, 67, and their high variability. However, taking into account the success at n metric, ALBP achieved a 82.00% retrieving 40 images, which means that 82 out of 100 images have at least one correspondence in the top 40 retrieved images. This is a very interesting result for the criminal police due to ease the task of evaluate a grid of 40 images at a glance in order to link images.

All in all, in this paper we introduced a new, efficient and powerful texture descriptor enhancer that adds statistical information about the gradient of the image. Based on the results obtained, we consider that CLOSIB can be regarded as a descriptor enhancer of wide purpose that, when it is fused with other descriptors, it provides new relevant information improving in this way the classification results. In the future, we will evaluate the performance obtained when combining CLOSIB with other different texture descriptors to determinate with which ones it works better and its limitations, if any.

Chapter 8

Conclusions and outlook

8.1. Work summary

In this dissertation, three new texture descriptors have been proposed.

Following a time line, the first one was ALBPS, which provides oriented standard deviation information to improve the description of the Local Binary Patterns. This algorithm was tested on two different application fields: material recognition and vitality classification of boar sperm heads.

Due to the high performance demonstrated in both cases, achieving higher hit rates than some state-of-the-art recent techniques developed in the last years, we proposed a second descriptor called Local Oriented Statistical Information Booster (LOSIB). LOSIB is a texture booster developed in order to enhance global and local descriptors. LOSIB extracts the local oriented information of the images taking into account the means of the gray value difference of the pixels and their neighbours along some given orientations. The concept of being presented as a booster allows to do an early fusion of LOSIB with any descriptor in order to include local oriented information of the image. We have used this booster to increase the performance of systems developed for material recognition, for acrosome integrity of boar spermatozoa and for tool wear recognition. In all the cases, LOSIB demonstrated a high increment of the classification performances adding just a few extra features with an additional low computational cost.

Finally, we proposed a generalisation of our previous method named CLOSIB from Complete LOSIB. CLOSIB highly improves the description power of LOSIB adding not only the first oriented statistical order information of the image gradient but also the second one. Furthermore, we presented three variants of CLOSIB: H-CLOSIB, useful for embedded systems or machines with a low RAM memoy; M-CLOSIB a multiscale descriptor which extracts information for consecutive neighborhoods and HM-CLOSIB, which is a multiscale version of H-CLOSIB. This descriptor is presented as a comprehensive mathematical formula. These methods have been evaluated for material recognition, and also two specific fields against crime: face recognition and textile retrieval. We can conclude that CLOSIB can be regarded as a descriptor enhancer of wide purpose that provides new relevant information about the statistical orientation of the image gradient.

8.2. General conclusions

The main conclusions extracted from this dissertation may be summarized as follows:

1. *A new method called Adaptive Local Binary Pattern with oriented Standard deviation (ALBPS) has been proposed. For material recognition it achieved a 61.47% of hit rate whereas for sperm vitality assessment it yielded a FScore of 0.886, both in combination with WCF13. In both cases it outperformed the results obtained using the literature descriptors.*

2. *A new booster for texture description called Local Oriented Statistical Information Booster (LOSIB) has been proposed demonstrating that when it is fused with another texture description, the results improve significantly.* A full evaluation has been performed using the most challenging datasets for material recognition and the most valuable descriptors in the state-of-the-art.

3. *100% of precision and 0.9975 of FScore has been achieved for the evaluation of the state of the acrosome in boar spermatozoa by combining LOSIB with Fourier information and Haralick features extracted on the Wavelet responses.*

4. *A generalised version of the LOSIB booster called Complete LOSIB (CLOSIB) has been presented. Experiments carried out have shown that CLOSIB improves the results achieved by LOSIB in combination with texture descriptors in multiple applications.*

5. *A new CBIR method for textile retrieval of fabrics like carpets, curtains, etc. has been presented.* This is a novel method that deals with an unexplored field in artificial vision and it combines clustering and description techniques for indoor scenes.

8.3. Future research lines

In future, we will evaluate the early fusion of CLOSIB with other different texture descriptors to determinate the best combinations and their limitations. Furthermore, we will explore the effectiveness of our booster in other fields of application where texture description is essential such as medical, pedestrian or object detection.

Another interesting study related with this dissertation is to evaluate the combination of CLOSIB booster with other type of descriptors different from texture features, such as shape and global or local invariant feature descriptors.

Finally, in order to increase the power of CLOSIB, we will implement and evaluate more statistical information metrics such as skew or kurtosis to evaluate the possibility of adding them into our booster.

Capítulo 9

Conclusiones y perspectivas

9.1. Resumen del trabajo

En esta tesis se propusieron tres descriptores de textura. Siguiendo una línea temporal, el primero fue ALBPS, que nos proporciona información sobre la desviación estándar orientada para mejorar al descriptor clásico LBP. Este algoritmo se probó en dos campos de aplicación diferentes: reconocimiento de materiales y clasificación de espermatozoides de verraco en función de su vitalidad. Debido al buen funcionamiento en ambas aplicaciones, propusimos un segundo descriptor llamado LOSIB. LOSIB es un potenciador de textura desarrollado para mejorar descriptores locales y globales. Extrae la información de las orientaciones locales teniendo en cuenta la media de las diferencias de los valores de gris de los píxeles y sus vecinos en diferentes orientaciones. El hecho de que sea presentado como un potenciador, nos permite hacer una fusión temprana de LOSIB con cualquier descriptor para incluit esa información de orientaciones locales de la imagen. Además, hemos usado el potenciador para mejorar la eficacia de los sistemas desarrollados para reconocimiento de materiales, clasificación de la integridad del acrosoma de espermatozoides y monitorización del desgaste de herramientas de corte. En todos estos casos, LOSIB demostró un incremento de la tasa de acierto en las clasificaciones añandiendo solo un pequeño número de características adicionales, de manera que sigue siendo un descriptor con bajo coste computacional. Finalmente, propusimos una generalización de LOSIB llamado CLOSIB (de Complete LOSIB). Este método incrementa el poder de descripción de LOSIB añadiendo no solo información sobre el orden del primer estadístico orientado de la imagen sino también del segundo. Además presentamos tres variantes de CLOSIB: H-CLOSIB, ideal para sistemas embebidos o máquinas con poca memoria RAM; M-CLOSIB, un descriptor multiescala que extrae información de vecinos consecutivos y HM-CLOSIB que es una versión multiescala de H-CLOSIB. Estos métodos han sido evaluados con el mismo conjunto de imágenes de recuperación de materiales, KTH TIPS2-a, y también en dos campos específicos en contra del crimen: reconocimiento de rostros y recuperación de textiles. Teniendo en cuenta todo esto, podemos concluir que CLOSIB puede ser considerado como un algoritmo para mejorar descritores y además con un amplio rango de aplicación, que nos da información sobre la orientación estadística del gradiente de la imagen.

9.2. Conclusiones generales

Las principales conclusiones que podemos extraer de esta tesis, se pueden resumir como sigue:

1. Se propuso un nuevo método llamado ALBPS (Adaptive Local Binary Pattern with oriented Standard deviation) basado en el presentado por Guo, Zhang, Zhang and Zhang (2010). Nuestra propuesta obtiene un 61.47 % de tasa acierto mejorando los métodos de reconocimiento de materiales de la literatura. También mejora la clasificación del acrosoma de espermatozoides alcanzando un FScore de 0.886 cuando se combina con WCF13.

2. Se propuso un nuevo potenciador de descriptores de textura llamado LOSIB (Local Oriented Statistical Information Booster). Se llevó a cabo una evaluación completa usando el conjunto de imágenes de recuperación de materiales que supone un mayor reto y los descriptores mejor valorados en el estado del arte.

3. Para evaluar el estado del acrosoma de espermatozoides de verraco, se realizó una fusión temprana de LOSIB con información de Fourier y Wavelet. Obtuvimos un 100 % de precisión y un FScore de 0.9975, resolviendo así por completo el problema de la evaluacion de acrosomas.

4. Se presentó una versión generalizada de LOSIB llamada CLOSIB (Complete LOSIB). Este nuevo método añade más posibilidades de configuración en relación con LOSIB. Los experimentos que llevamos a cabo mostraron que CLOSIB es una buena herramienta para mejorar los resultados de otros descriptores de textura en múltiples aplicaciones.

5. Se presentó un nuevo método de recuperación de imagen basado en contenido para recuperar tejidos como alfombras, cortinas, etc. Es un método nuevo que trata un campo inexplorado dentro de la visión artificial y combina técnicas de agrupamientos y de descripción para escenas de interior.

9.3. Líneas futuras

En el futuro, evaluaremos la fusión temprana de CLOSIB con otros descriptores de textura diferentes para determinar las mejores combinaciones y sus limitaciones. Además, exploraremos la eficacia de nuestro potenciador en otros campos de aplicación donde la descripción de la textura es esencial, como por ejemplo la detección médica, o reconocimiento de peatones o de objetos. Otro estudio interesante relacionado con esta tesis es evaluar la combinación de CLOSIB con otro tipo de descriptores que no sean de textura, como por ejemplo, de forma o de características invariantes locales y globales. Finalmente, con el fin de aumentar el potencial de CLOSIB, también sería interesante estudiar más información estadística como sesgo o curtosis para evaluar la posibilidad de incluirlos en nuestro potenciador.

Bibliography

Ahmadvand, A. and Daliri, M. R.: 2016, Rotation invariant texture classification using extended wavelet channel combining and ll channel filter bank, *Knowledge-Based Systems* **97**, 75 – 88.

Alegre, E., Bieh, 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 Sanchez, 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., Garcia-Olalla, O., Gonzalez-Castro, V. and Joshi, S.: 2011, Boar spermatozoa classification using longitudinal and transversal profiles (LTP) descriptor in digital images., *International Workshop on Combinatorial Image Analysis*.

Alegre, E., Garcia-Ordas, M., Gonzalez-Castro, V. and Karthikeyan, S.: 2011, Vitality assessment of boar sperm using NCSR texture descriptor in digital images, *Iberian Conference on Pattern Recognition and Image Analysis proceedings*.

Alegre, E., Gonzalez-Castro, V., Alaiz-Rodriguez, R. and Garcia-Ordas, M. T.: 2012, Texture and moments-based classification of the acrosome integrity of boar spermatozoa images, *Computer Methods and Programs in Biomedicine* .

Alegre, E., Gonzalez-Castro, V., Suarez, S. and Castejon-Limas, M.: 2009, Comparison of supervised and unsupervised methods to classify boar acrosomes using texture descriptors, *51st International Symposium ELMAR*.

Alegre, E., González-Castro, V., Suárez, S. and Castejón, M.: 2009, Comparison of supervised and unsupervised methods to classify boar acrosomes using texture descriptors, *Proceedings ELMAR-2009*, Vol. 1, pp. 65–70.

Anastassopoulos, V. and Venetsanopoulos, A.: 1991, The classification properties of the pecstrum and its use for pattern identification, *Circuits, Systems and Signal Processing, 10(3), 293-326.* .

Aptoula, E. and Lefevre, S.: 2007, On morphological color texture characterization, *International symposium on mathematical morpholog* .

Aptoula, E. and Lefevre, S.: 2011, Morphological texture description of grey scale and color images, *in* P. W. Hawkes (ed.), *Advances in Imaging and Electron Physics*, Vol. 169 of *Advances in Imaging and Electron Physics*, Elsevier, pp. 1–74.

Arivazhagan, S. and Ganesan, L.: 2003, Texture segmentation using wavelet transform, *Pattern Recognition Letters* **24**(16), 3197 – 3203.

Ayala, G. and Domingo, J.: 2001, Spatial size distribution: applications to shape and texture analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligenc* .

Baeg, S., Batman, S., Dougherty, E. R., Kamat, V., Kehtarnavaz, N. and Kim, S.: 1999, Unsupervised morphological granulometric texture segmentation of digital mammograms, *Journal of Electronic Imaging* .

Bagdanov, A. and Worring, M.: 2002, Granulometric analysis of document images, *In International Conference on Pattern Recognition* .

Bakhshipour, A., Jafari, A., Nassiri, S. M. and Zare, D.: 2017, Weed segmentation using texture features extracted from wavelet sub-images, *Biosystems Engineering* **157**, 1 – 12.

Bangham, J. A., Ling, P. D. and Harvey, R.: 1996, Scale-space from nonlinear filters, *IEEE Transactions on Pattern Analysis and Machine Intelligenc* .

Barnich, O., Jodogne, S. and Van Droogenbroeck, M.: 2006, Robust analysis of silhouettes by morphological size distributions, *In Advanced Concepts for Intelligent Vision Systems* .

Bashar, M., Ohnishi, N., Matsumoto, T., Takeuchi, Y., Kudo, H. and Agusa, K.: 2005, Image retrieval by pattern categorization using wavelet domain perceptual features with {LVQ} neural network, *Pattern Recognition Letters* **26**(15), 2315 – 2335.

Bhuiyan, M. and Choudhury, I.: 2014, Review of sensor applications in tool condition monitoring in machining, *Comprehensive Materials Processing* **13**, 539–569.

Bijar, A., Mikaeili, M., Benavent, A. and Khayati, R.: 2012, Segmentation of sperm's acrosome, nucleus and mid-piece in microscopic images of stained human semen smear, *Communication Systems, Networks Digital Signal Processing (CSNDSP), 2012 8th International Symposium on*, pp. 1–6.

Brilakis, I. K., Soibelman, L. and Shinagawa, Y.: 2006, Construction site image retrieval based on material cluster recognition, *Advanced Engineering Informatics* **20**(4), 443 – 452.

Bugatti, P. H., Kaster, D. S., Ponciano-Silva, M., Jr., C. T., Azevedo-Marques, P. M. and Traina, A. J.: 2014, Prosper: Perceptual similarity queries in medical {CBIR} systems through user profiles, *Computers in Biology and Medicine* **45**, 8 – 19.

Caputo, B., Hayman, E. and Mallikarjuna, P.: 2005, Class-specific material categorisation, *ICCV*.

Carbunaru, A. E., Coltuc, D., Jourlin, M. and Frangu, L.: 2009, A texture descriptor for textile image retrieval, *2009 International Symposium on Signals, Circuits and Systems*, pp. 1–4.

Castrillon-Santana, M., Lorenzo-Navarro, J. and Ramon-Balmaseda, E.: 2015, Fusion of holistic and part based features for gender classification in the wild, *in* V. Murino, E. Puppo, D. Sona, M. Cristani and C. Sansone (eds), *New Trends in Image Analysis and Processing – ICIAP 2015 Workshops*, Vol. 9281 of *Lecture Notes in Computer Science*, Springer International Publishing, pp. 43–50.

Chaddad, A., Tanougast, C., Dandache, A. and Bouridane, A.: 2011, Extraction of haralick features from segmented texture multispectral bio-images for detection of colon cancer cells, *Informatics and Computational Intelligence (ICI), 2011 First International Conference on*, pp. 55 –59.

Chan, C.-H., Kittler, J. and Messer, K.: 2007, *Multi-scale Local Binary Pattern Histograms for Face Recognition*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 809–818.

Chang, L., Duarte, M. M., Sucar, L. and Morales, E. F.: 2012, A bayesian approach for object classification based on clusters of {SIFT} local features, *Expert Systems with Applications* **39**(2), 1679 – 1686.

Chee-Way Chonga, P. Raveendran, R. M.: 2004, Translation and scale invariants of legendre moments, *Pattern Recognition* **37**(1), 119 – 129.

Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X. and Gao, W.: 2010, Wld: A robust local image descriptor, *PAMI*.

Chen, J., Tian, J., Lee, N., Zheng, J., Smith, R. and Laine, A.: 2010, A partial intensity invariant feature descriptor for multimodal retinal image registration, *Biomedical Engineering, IEEE Transactions on* **57**(7), 1707 –1718.

Chen, L.-C., Hsieh, J.-W., Yan, Y. and Chen, D.-Y.: 2015, Vehicle make and model recognition using sparse representation and symmetrical {SURFs}, *Pattern Recognition* **48**(6), 1979 – 1998.

Chun, J. C. and Kim, W. G.: 2013, Textile image retrieval using composite feature vectors of color and wavelet transformed textural property, *Measurement Technology and Engineering Researches in Industry*, Vol. 333 of *Applied Mechanics and Materials*, Trans Tech Publications, pp. 822–827.

Craswell, N.: 2009, *Success at n*, Springer US, Boston, MA, pp. 2875–2876.

Dalal, N. and Triggs, B.: 2005, 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.

de Araujo, S. A., Pessota, J. H. and Kim, H. Y.: 2015, Beans quality inspection using correlation-based granulometry, *Engineering Applications of Artificial Intelligence* **40**, 84 – 94.

Dharmagunawardhana, C., Mahmoodi, S., Bennett, M. and Niranjan, M.: 2016, Rotation invariant texture descriptors based on gaussian markov random fields for classification, *Pattern Recognition Letters* **69**, 15 – 21.

Dornaika, F., Bosaghzadeh, A., Salmane, H. and Ruichek, Y.: 2014, A graph construction method using {LBP} self-representativeness for outdoor object categorization, *Engineering Applications of Artificial Intelligence* **36**(0), 294 – 302.

Dougherty, E. R. and Astola, J. T.: 1994, An introduction to nonlinear image processing., *n Tutorial texts in optical engineerin* .

Drbohlav, O. and Leonardis, A.: 2010, Towards correct and informative evaluation methodology for texture classification under varying viewpoint and illumination, *Computer Vision and Image Understanding* **114**(4), 439 – 449. Special issue on Image and Video Retrieval Evaluation.

Elaiwat, S., Bennamoun, M., Boussaid, F. and El-Sallam, A.: 2015, A curvelet-based approach for textured 3d face recognition, *Pattern Recognition* **48**(4), 1235 – 1246.

Faria, A. V., Oishi, K., Yoshida, S., Hillis, A., Miller, M. I. and Mori, S.: 2015, Content-based image retrieval for brain mri: An image-searching engine and population-based analysis to utilize past clinical data for future diagnosis, *NeuroImage: Clinical* **7**, 367 – 376.

Farrell, F., Presicce, G., Brockett, C. and Foote, R.: 1998, Quantification of bull sperm characteristics measured by computer-assisted sperm analysis (CASA) and the relationship to fertility., *Theriogenology* **49**(4), 871–879.

Fayad, R.: 2010, Tool monitoring for drilling process applying enhanced neural networks., *The 2nd International Conference on Computer and Automation Engineering (ICCAE)*.

Fazeli, A., Hage, W. ., Cheng, F.-P., F.Voorhout, W., Marks, A., Bevers, M. M. and Colenbrander, B.: 1997, Acrosome-lntact boar spermatozoa initiate binding to the homologous zona pellucida invitro, *Biology of reproduction* **56**, 430–438.

Feng, L., Bhanu, B. and Heraty, J.: 2016, A software system for automated identification and retrieval of moth images based on wing attributes, *Pattern Recognition* **51**, 225 – 241.

Fernández-Abia, A., Barreiro, J., de Lacalle, L. L. and González-Madruga, D.: 2014, Effect of mechanical pre-treatments in the behaviour of nanostructured pvd-coated tools in turning, *Int J Advanced Manufacturing Technology* **On-line acess**.

Florindo, J. and Bruno, O.: 2016, Texture analysis by fractal descriptors over the wavelet domain using a best basis decomposition, *Physica A: Statistical Mechanics and its Applications* **444**, 415 – 427.

Gadre, V. and Patney, R.: 1992, Multiparametric multiscale filtering, multiparametric granulometries and the associated pattern spectra, *In International Symposium on Circuits and Systems* .

Gao, H., Li, D., Xu, M., Zhao, M., Shi, X. and Huang, H.: 2010, Tool wear monitoring based on novel evolutionary artificial neural networks, *Sixth International Conference on Natural Computation (ICNC)*.

Garcia-Olalla, O., García-Ordás, M., García-Ordas, D., Fernández-Robles, L. and Alegre, E.: 2012, Vitality assessment of boar sperm using n concentric squares resized and local binary pattern in gray scale images., *XXXIII Jornadas de Automatica*.

Ghosh, P. and Chanda, B.: 1998, Bi-variate pattern spectrum, *In International Symposium on Computer Graphics, Image Processing and Vision* .

González-Castro, V., Aláiz-Rodríguez, R. and Alegre, E.: 2012, Class distribution estimation based on the Hellinger distance, *Information Sciences* .

Gonzalez-Castro, V., Alegre, E., Morala-Arguello, P. and Suarez, S.: 2009, A combined and intelligent new segmentation method for boar semen based on thresholding and watershed transform, *International Journal of Imaging* **2**, 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*, Taylor and Francis Group London, pp. 165–168.

Goyal, K. and Singhai, J.: 2016, Texture-based self-adaptive moving object detection technique for complex scenes, *Computers & Electrical Engineering* .

Guo, Z., Zhang, L. and Zang, D.: 2010, A completed modeling of local binary pattern operator for texture classification, *Image Processing, IEEE Transactions on* .

Guo, Z., Zhang, L. and Zhang, D.: 2010, Rotation invariant texture classification using lbp variance (lbpv) with global matching, *Pattern Recognition* **43**(3), 706–719.

Guo, Z., Zhang, L., Zhang, D. and Zhang, S.: 2010, Rotation invariant texture classification using adaptive lbp with directional statistical features, *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pp. 285 –288.

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

Hafiane, A., Palaniappan, K. and Seetharaman, G.: 2015, Joint adaptive median binary patterns for texture classification, *Pattern Recognition* **48**(8), 2609 – 2620.

Hanbury, A., Kandaswamy, U. and Adjeroh, D. A.: 2005, Illumination-invariant morphological texture classification., *Mathematical morphology: 40 years on* .

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

Hegenbart, S., Uhl, A., Vecsei, A. and Wimmer, G.: 2013, Scale invariant texture descriptors for classifying celiac disease, *Medical Image Analysis* **17**(4), 458 – 474.

Heijmans, H. J. A. M., Nacken, P., Toet, A. and Vincent, L.: 1992, Graph morphology, *Journal of Visual Communication and Image Representation* .

Hirano, Y., Shibahara, H., Obara, H., Suzuki, T., Takamizawa, S., Yamaguchi, C., Tsunoda, H. and Sato, I.: 2001, Relationships between sperm motility characteristics assessed by the computer-aided sperm analysis (CASA) and fertilization rates in vitro., *J Assist Reprod Genet* **18**(4), 213–8.

Holmquist, J., Bengtsson, E., Eriksson, O., Nordin, B. and Stenkvist, B.: 1978, Computer analysis of cervical cells. automatic feature extraction and classification., *J Histochem Cytochem* **26**(11), 1000–1017.

Hsu, C.-W. and Lin, C.-J.: 2002, A comparison of methods for multiclass support vector machines, *IEEE Transactions on Neural Networks* **13**(2), 415–425.

Hu, M.-K.: 1962, Visual pattern recognition by moment invariants, *Information Theory, IRE Transactions on* **8**(2), 179–187.

Huang, Y.-F. and Lin, S.-M.: 2014, *Searching Images in a Textile Image Database*, Springer International Publishing, Cham.

Hussain, S. and Triggs, B.: 2012, Visual recognition using local quantized patterns, *in* A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato and C. Schmid (eds), *Computer Vision ECCV 2012*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 716–729.

Iqbal, K., Odetayo, M. O. and James, A.: 2012, Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics, *Journal of Computer and System Sciences* **78**(4), 1258 – 1277.

Jia, X., Yang, X., Cao, K., Zang, Y., Zhang, N., Dai, R., Zhu, X. and Tian, J.: 2014, Multi-scale local binary pattern with filters for spoof fingerprint detection, *Information Sciences* **268**, 91 – 102. New Sensing and Processing Technologies for Hand-based Biometrics Authentication.

Jiang, Y., Zheng, S., Yan, C. and Zhang, W.: 2009, The application of computer-assisted semen analysis parameters in semen quality analysis, *BioMedical Information Engineering, 2009. FBIE 2009. International Conference on Future*, pp. 256–258.

Kashyap, R. and Khotanzed, A.: 1986, A model-based method for rotation invariant texture classification, *IEEE Transactions on PAMI* **8**, 472–481.

Kavitha, M. S., An, S.-Y., An, C.-H., Huh, K.-H., Yi, W.-J., Heo, M.-S., Lee, S.-S. and Choi, S.-C.: 2015, Texture analysis of mandibular cortical bone on digital dental panoramic radiographs for the diagnosis of osteoporosis in korean women, *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology* **119**(3), 346 – 356.

Klimowicz-Bodys, M., Batkowski, F., Ochrem, A. and MA, M. S.: 2012, Comparison of assessment of pigeon sperm viability by contrast-phase microscope (eosin-nigrosin staining) and flow cytometry (sybr-14/propidium iodide (pi) staining), *Theriogenology* **77**(3), 628–35.

Kopac, J.: 1998, Influence of cutting material and coating on tool quality and tool life, *Journal of Materials Processing Technology* **78**(1-3), 95 – 103.

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

Lam, W. K. and Li, C. K.: 1996, Scale invariant texture classification by iterative morphological decomposition, *Electronics Letters* .

Lam, W. K. and Li, C. K.: 1997, Rotated texture classification by improved iterative morphological decomposition, *Vision, Image and Signal Processing* .

Lefevre, S., Weber, J. and Sheeren, D.: 2007, Automatic building extraction in vhr images using advanced morphological operators, *In IEEE Joint Workshop on Remote Sensing and Data Fusion over Urban Areas. Paris, France.* .

Li, H., Liu, Z., Huang, Y. and Shi, Y.: 2015, Quaternion generic fourier descriptor for color object recognition, *Pattern Recognition* **48**(12), 3895 – 3903.

Li, W. and Fritz, M.: 2012, Recognizing materials from virtual examples, *Proceedings of the 12th European Conference on Computer Vision - Volume Part IV*, ECCV'12, Springer-Verlag, Berlin, Heidelberg, pp. 345–358.

Li, W., Fu, P. and Cao, W.: 2010, Tool wear states recognition based on genetic algorithm and back propagation neural network model, *International Conference on Computer Application and System Modeling (ICCASM)*.

Li, W., Hease-Coat, V. and Ronsin, J.: 1996, Robust morphological features for texture classificatio, *In International Conference on Image Processing (Vol. 3, pp. 173-176). Lausanne , Switzerland.* .

Liang, C.-W. and Juang, C.-F.: 2015, Moving object classification using local shape and {HOG} features in wavelet-transformed space with hierarchical {SVM} classifiers, *Applied Soft Computing* **28**(0), 483 – 497.

Lim, G.: 1995, Tool-wear monitoring in machine turning, *Journal of Materials Processing Technology* **51**(1-4), 25 – 36.

Lin, T.-W. and Chou, Y.-F.: 2003, A comparative study of zernike moments, *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, pp. 516–519.

Liu, F., Tang, Z. and Tang, J.: 2013, Wlbp: Weber local binary pattern for local image description, *Neurocomputing* **120**(0), 325 – 335. Image Feature Detection and Description.

Liu, P., Guo, J.-M., Chamnongthai, K. and Prasetyo, H.: 2017, Fusion of color histogram and lbp-based features for texture image retrieval and classification, *Information Sciences* **390**, 95 – 111.

Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J.: 1998, Coding facial expressions with gabor wavelets, *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 200–205.

Mahersia, H. and Hamrouni, K.: 2015, Using multiple steerable filters and bayesian regularization for facial expression recognition, *Engineering Applications of Artificial Intelligence* **38**(0), 190 – 202.

Mallik, J., Samal, A. and Gardner, S. L.: 2010, A content based image retrieval system for a biological specimen collection, *Computer Vision and Image Understanding* **114**(7), 745 – 757.

Manjunath, B. S. and Ma, W. Y.: 1996, Texture features for browsing and retrieval of image data, *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(8), 837–842.

Maragos, P.: 1989, Pattern spectrum and multiscale shape representation., *IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(7), 701-716.* .

Marsico, M. D., Nappi, M., Riccio, D. and Wechsler, H.: 2015, Robust face recognition after plastic surgery using region-based approaches, *Pattern Recognition* **48**(4), 1261 – 1276.

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.

Nanni, L., Brahnam, S. and Lumini, A.: 2012, Random interest regions for object recognition based on texture descriptors and bag of features, *Expert Systems with Applications* **39**(1), 973 – 977.

Ojala, T., Pietikainen, M. and Harwood, D.: 1994, Performance evaluation of texture measures with classification based on kullback discrimination of distributions, *Proceedings of the 12th IAPR International Conference on Pattern Recognition (ICPR 1994)*.

Ojala, T., Pietikainen, M. and Maenpaa, T.: 2002, Multiresolution gray-scale and rotation invariant texture classification with Local Binary Patterns, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(7), 971–987.

Otsu, N.: 1979, A threshold selection method from gray-level histograms, *IEEE Transactions on Systems, Man and Cybernetics* **9**(1), 62–66.

Painuli, S., Elangovan, M. and Sugumarana, V.: 2014, Tool condition monitoring using k-star algorithm, *Expert Systems with Applications* **41**, 2638–2643.

Pan, Z., Li, Z., Fan, H. and Wu, X.: 2017, Feature based local binary pattern for rotation invariant texture classification, *Expert Systems with Applications* pp. –.

Pascual-Gaspar, J., Olmedo, H., Exposito, A., Exposito, A. and Finat, J.: 2008, A simple and effective system for computer-assisted semen analysis, *Advances in Medical, Signal and Information Processing, 2008. MEDSIP 2008. 4th IET International Conference on*, pp. 1–4.

Patel, B., Maheshwari, R. and Balasubramanian, R.: 2016, Multi-quantized local binary patterns for facial gender classification, *Computers & Electrical Engineering* **54**, 271 – 284.

Petkov, N., Alegre, E., Biehl, M. and Sanchez, L.: 2007, LVQ acrosome integrity assessment of boar sperm cells, *in* J. M. R. S. Tavares and R. M. N. Jorge (eds), *Computational Modeling of Objects Represented in Images, Proc. CompImage Symposium 2006*, Taylor and Francis, pp. 337–342.

Pratama, M., Er, M., Li, X., Gan, O., Oentaryo, R., Linn, S., Zhai, L. and Arifin, I.: 2011, Tool wear prediction using evolutionary dynamic fuzzy neural (edfnn) network. 7-10 nov., *37th Annual Conference on IEEE Industrial Electronics Society IECON* .

Qian, Y., Tian, J., Liu, L., Zhang, Y. and Chen., Y.: 2010, A tool wear predictive model based on svm, *Chinese Control and Decision Conference (CCDC)*.

Rakvongthai, Y. and Oraintara, S.: 2013, Statistical texture retrieval in noise using complex wavelets, *Signal Processing: Image Communication* **28**(10), 1494 – 1505.

Ribaric, S. and Lopar, M.: 2012, Palmprint recognition based on local haralick features, *Electrotechnical Conference (MELECON), 2012 16th IEEE Mediterranean*, pp. 657–660.

Rodenacker, K. and Bengtsson, E.: 2003, A feature set for cytometry on digitized microscopic images, *Anal Cell Pathol* **25**(1), 1–36.

Ruggeri, A. and Pajaro, S.: 2002, Automatic recognition of cell layers in corneal confocal microscopy images, *Computer Methods and Programs in Biomedicine* **68**(1), 25 – 35.

Sanchez, L., Petkov, N. and Alegre, E.: 2005, Classification of boar spermatozoid head images using a model intracellular density distribution, *Progress in Pattern Recognition, Image Analysis and Applications: Proceedings of the 10th Iberoamerican Congress on Pattern Recognition*.

Sanchez, 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**, 38–43.

Satpathy, A., Jiang, X. and Eng, H.-L.: 2014, Lbp-based edge-texture features for object recognition, *Image Processing, IEEE Transactions on* **23**(5), 1953–1964.

Serra, J.: 1982, Image analysis and mathematical morphology, *Academic Press* .

Sharma, G., Ul Hussain, S. and Jurie, F.: 2012, Local higher-order statistics (lhs) for texture categorization and facial analysis, *ECCV - European Conference on Computer Vision*, Florence, Italie.

Shen, J., Zuo, X., Li, J., Yang, W. and Ling, H.: 2017, A novel pixel neighborhood differential statistic feature for pedestrian and face detection, *Pattern Recognition* **63**, 127 – 138.

Shenoy, P. D., Iyengar, S. S., Raja, K. B., R, V. K., Buyya, R., Patnaik, L. M., Rangaswamy, Y., Raja, K. and Venugopal, K.: 2015, FRDF: Face recognition using fusion of DTCWT and FFT features, *Procedia Computer Science* **54**, 809 – 817.

Shih, H.-C. and Yu, K.-C.: 2015, {SPiraL} aggregation map (splam): A new descriptor for robust template matching with fast algorithm, *Pattern Recognition* **48**(5), 1707 – 1723.

Shotton, J. D. J., Winn, J., Rother, C. and Criminisi, A.: 2009, Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context., *International Journal of Computer Vision* .

Siemieniuch, M. and Woclawek-Potocka, I.: 2008, Assessment of selected quality parameters of epididymal cat (felis catus s. domestica, l. 1758) sperm using flow cytometry method and computer assisted sperm analyser., *Reprod Domest Anim* **43**(5), 633–637.

Soille, P.: 2003, *Morphological Image Analysis: Principles and Applications*, 2 edn, Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Son, H., Kim, C., Hwang, N., Kim, C. and Kang, Y.: 2014, Classification of major construction materials in construction environments using ensemble classifiers, *Advanced Engineering Informatics* **28**(1), 1 – 10.

Southam, P. and Harvey, R.: 2004, Compact rotation-invariant texture classification., *In International Conference on Image Processing* .

Southam, P. and Harvey, R.: 2005a, Texture granularities, *In International Conference on Image Analysis and Processing* .

Southam, P. and Harvey, R.: 2005b, Towards texture classification in real scenes, *In British Machine Vision Conference* .

Southam, P. and Harvey, R. W.: 2009, Texture classification via morphological scale-space: Tex-mex features, *Journal of Electronic Imaging* .

Srinivas, M., Naidu, R. R., Sastry, C. and Mohan, C. K.: 2015, Content based medical image retrieval using dictionary learning, *Neurocomputing* **168**, 880 – 895.

Tan, M., Hu, Z., Wang, B., Zhao, J. and Wang, Y.: 2016, Robust object recognition via weakly supervised metric and template learning, *Neurocomputing* **181**, 96 – 107. Big Data Driven Intelligent Transportation Systems.

Teti, R.: 2002, Machining of composite materials, *{CIRP} Annals - Manufacturing Technology* **51**(2), 611 – 634.

Tuceryan, M. and Jain, A. K.: 1998, Texture analysis, *The handbook of pattern recognition and computer vision* .

Valcárcel, A., de las Heras, M., Pérez, L., Moses, D. and Baldassarre, H.: 1997, Assessment of the acrosomal status of membrane-intact ram spermatozoa after freezing and thawing, by simultaneous lectin/hoechst 33258 staining., *Animal Reproduction Science* **45**, 299–309.

Valkealahti, K. and Oja, E.: 1998, Reduced multidimensional co-occurrence histograms in texture classification, *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **20**(1), 90–94.

Vantman, D., Koukoulis, G., Dennison, L., Zinaman, M. and Sherins, R.: 1988, Computer-assisted semen analysis: evaluation of method and assessment of the influence of sperm concentration on linear velocity determination., *Fertil Steril*, p. 510?515.

Viola, P. and Jones, M.: 2001, Rapid object detection using a boosted cascade of simple features, *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, Vol. 1, IEEE, pp. I–511.

Wan, W., Zhou, Z., Zhao, J. and Cao, F.: 2015, A novel face recognition method: Using random weight networks and quasi-singular value decomposition, *Neurocomputing* **151, Part 3**, 1180 – 1186.

Wang, D., Haese-Coat, V., Bruno, A. and Ronsin, J.: 1993, Texture classification and segmentation based on iterative morphological decomposition, *Journal of Visual Communication and Image Representation, 4(3), 197-214.* .

Wang, G., Yang, Y., Zhang, Y. and Xie, Q.: 2014, Vibration sensor based tool condition monitoring using support vector machine and locality preserving projection, *Sensors and Actuators A:Physical* **209**, 24–32.

Wang, K., Bichot, C.-E., Li, Y. and Li, B.: 2017, Local Binary Circumferential and radial derivative pattern for texture classification, *Pattern Recognition* .

Wang, Y., Jia, Y., Yuchi, M. and Ding, M.: 2011, The computer-assisted sperm analysis (CASA) technique for sperm morphology evaluation, *Intelligent Computation and Bio-Medical Instrumentation (ICBMI), 2011 International Conference on*, pp. 279–282.

Weckenmann, A. and Nalbantic, K.: 2003, Precision measurement of cutting tools with two matched optical 3d-sensors, *{CIRP} Annals - Manufacturing Technology* **52**(1), 443 – 446.

Wen, Z., Li, Z., Peng, Y. and Ying, S.: 2016, Virus image classification using multi-scale completed local binary pattern features extracted from filtered images by multi-scale principal component analysis, *Pattern Recognition Letters* **79**, 25 – 30.

Wilkinson, M. H. F.: 2002, Generalized pattern spectra sensitive to spatial information, *In International Conference on Pattern Recognition* .

Yang, B., Yu, H. and Hu, R.: 2015, Unsupervised regions based segmentation using object discovery, *Journal of Visual Communication and Image Representation* **31**, 125 – 137.

Zang, F., Zhang, J. and Pan, J.: 2012, Face recognition using elasticfaces, *Pattern Recognition* **45**(11), 3866 – 3876.

Zhang, D. and Lu, G.: 2003, A comparative study on shape retrieval using fourier descriptors with different shape signatures, *Journal of Visual Communication and Image Representation* **1**(14 (1)), 41–60.

Zhang, J., Zhao, H. and Liang, J.: 2013, Continuous rotation invariant local descriptors for texton dictionary-based texture classification, *Computer Vision and Image Understanding* **117**(1), 56 – 75.

Zhang, L., Chu, R., Xiang, S., Liao, S. and Li, S. Z.: 2007, *Face Detection Based on Multi-Block LBP Representation*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 11–18.

Zhang, Q., Lin, J., Tao, Y., Li, W. and Shi, Y.: 2017, Salient object detection via color and texture cues, *Neurocomputing* **243**, 35 – 48.

Zheng, C.-H., Pei, W.-J., Yan, Q. and Chong, Y.-W.: 2017, Pedestrian detection based on gradient and texture feature integration, *Neurocomputing* **228**, 71 – 78. Advanced Intelligent Computing: Theory and Applications.

Zheng, Y. and Sarem, M.: 2016, A fast region segmentation algorithm on compressed gray images using non-symmetry and anti-packing model and extended shading representation, *Journal of Visual Communication and Image Representation* **34**, 153 – 166.

Zhu, J., Yu, J., Wang, C. and Li, F.-Z.: 2015, Object recognition via contextual color attention, *Journal of Visual Communication and Image Representation* **27**, 44 – 56.

Zhu, Z. and Brilakis, I.: 2010, Parameter optimization for automated concrete detection in image data, *Automation in Construction* **19**(7), 944 – 953.

Annex A: Research activities

Publications related with this manuscript

- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, Eduardo Fidalgo. "Textile retrieval using MSER region detection and LBP variants." Computer vision and Image Understanding. Not submitted.
- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, Manuel Castejón Limas. "Boosting Texture-Based Classification Using a Gradient Statistical Information Descriptor" Transactions on Image Processing. Under consideration. 2017.
- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, Patrik Malm, Ewert Bengtsson, "Acrosome integrity assessment of boar spermatozoa images using an early fusion of texture and contour descriptors", Computer Methods and Programs in Biomedicine, 2015.
- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, Víctor González-Castro, "Local Oriented Statistical Information Booster (LOSIB) for texture retrieval", ICPR International Conference on Pattern Recognition, Suecia, 2014.
- Oscar García-Olalla Olivera, Enrique Alegre, Joaquín Barreiro, Laura Fernández-Robles, María Teresa García-Ordás, "Tool wear classification using texture descriptors based on Local Binary Pattern", XXXV Jornadas de Automática, Valencia, 2014.
- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, María Teresa García-Ordás, Diego García-Ordás, "Adaptive local binary pattern with oriented standard deviation (ALBPS) for texture classification", EURASIP Journal on Image and Video Processing, 2013.
- Oscar García-Olalla Olivera, Enrique Alegre, María Teresa García-Ordás, Laura Fernández-Robles, "Evaluation of LBP Variants Using Several Metrics and kNN Classifiers", Similarity Search and applications, A Coruña, 2013.
- Oscar García-Olalla Olivera, María Teresa García-Ordás, Laura Fernández-Robles, Diego García Ordás, Enrique Alegre, "Comparison of different Local Binary Pattern variants for material recognition using KTH-TIPS 2a dataset", XXXIV Jornadas de Automática, Terrasa, 2013.
- Oscar García-Olalla Olivera, Enrique Alegre, Laura Fernández-Robles, María Teresa García-Ordás, "Vitality assessment of boar sperm using an adaptive LBP based on oriented deviation", Asian Conference in Computer Vision Workshop, Corea del sur, 2012.

- Oscar García-Olalla Olivera, María Teresa García-Ordás, Laura Fernández-Robles, Enrique Alegre, ”Vitality assessment of boar sperm using N Concentric Squares resized and Local binary pattern in gray scale images”, XXXIII Jornadas de automática, Vigo 2012.

Other publications

- María Teresa García-Ordás, Enrique Alegre, Oscar García-Olalla Olivera, Diego García-Ordás, ”: Evaluation of Different Metrics for Shape Based Image Retrieval Using a New Contour Points Descriptor”, International Conference on Similarity Search and Applications (SISAP), 2013.
- Laura Fernández-Robles, Enrique Alegre, María Teresa García-Ordás, Oscar García-Olalla Olivera, Diego García-Ordás, Eduardo Fidalgo Fernández, ”Combining SURF with global texture descriptors for classifying boar sperm”, XXXIV Jornadas de automática, Terrasa 2013.
- Eduardo Fidalgo Fernández, Laura Fernández-Robles, María Teresa García-Ordás, Oscar García-Olalla Olivera, Enrique Alegre, ”: Evaluation of shape and color descriptors by using bag of Word techniques with one vs all classification”, XXXIV Jornadas de automática, Terrasa 2013.
- Diego García-Ordás, Enrique Alegre, María Teresa García-Ordás, Oscar García-Olalla Olivera, Laura Fernández-Robles, ”Robustness to rotation in perceptual hashing methods via dominant orientation”, XXXIV Jornadas de automática, Terrasa 2013.
- María Teresa García-Ordás, Oscar García-Olalla Olivera, Laura Fernández-Robles, Diego García Ordás, Enrique Alegre, ”: Rotation Invariant Contour Points Descriptor Histogram for shape based image retrieval”, XXXIV Jornadas de automática, Terrasa 2013.
- Diego García-Ordás, Laura Fernández-Robles, Enrique Alegre, María Teresa García-Ordás, Oscar García-Olalla Olivera, ”: Automatic tampering detection in spliced images with different compression levels”, 6th Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA), Madeira (Portugal), 2013.
- María Teresa García-Ordás, Laura Fernández-Robles, Oscar García-Olalla Olivera, Diego García-Ordás, Enrique Alegre, ”: Boar spermatozoa classification using local invariant features and bag of words”, XXXIII Jornadas de automática, Vigo 2012.
- Laura Fernández-Robles, Oscar García-Olalla Olivera, María Teresa García Ordás, Diego García-Ordás, Enrique Alegre, ”: SVM approach to classify boar acrosome integrity of a multi-features SURF description”, XXXIII Jornadas de automática, Vigo 2012.
- Víctor González-Castro, Enrique Alegre, Oscar García-Olalla Olivera, Diego García-Ordás, María Teresa García-Ordás, Laura Fernández-Robles, ”: Curvelet-Based Texture Description to Classify Intact and Damaged Boar Spermatozoa”, : 9th International Conference (ICIAR), Aveiro (Portugal), 2012.

- Sir Alexci Suárez-Castrillón, Joaquín Barreiro, Enrique Alegre, María Teresa García-Ordás, Oscar García-Olalla Olivera, ”Material surface characterization using Laws descriptors”, DAAAM International Symposium, Vienna (Austria), 2011.
- Laura Fernández-Robles, Víctor González-Castro, Oscar García-Olalla Olivera, María Teresa García-Ordás, Enrique 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), Olhão (Portugal), 2011.
- Patricia Morala Argüello, Joaquín Barreiro, Enrique Alegre, María Teresa García-Ordás, Oscar García-Olalla Olivera, Daniel González Madruga, ”Surface roughness prediction from combination of cutting forces, turning vibrations and machining conditions using artificial neural networks”, IV Manufacturing Engineering Society International Congress (CSIF-MESIC), Cadiz 2011.
- Laura Fernández-Robles, María Teresa García-Ordás, Diego García-Ordás, Oscar García-Olalla Olivera, Enrique Alegre, ”: Acrosome evaluation of spermatozoa cells using SIFT and classical texture Descriptors”, XXXII Jornadas de automática, Sevilla 2011.
- Diego García-Ordás, Oscar García-Olalla Olivera, Laura Fernández-Robles, María Teresa García-Ordás, Enrique Alegre, ”: Video segmentation combining depth maps and intensity images”, XXXII Jornadas de automática, Sevilla 2011.
- María Teresa García-Ordás, Laura Fernández-Robles, Oscar García-Olalla Olivera, Diego García-Ordás, Enrique Alegre, ”Words recognition using methods of word shape coding”, XXXII Jornadas de automática, Sevilla 2011.
- Enrique Alegre, Oscar García-Olalla Olivera, Víctor González-Castro, Swapna Joshi, ”Boar Spermatozoa Classification Using Longitudinal and Transversal Profiles (LTP) Descriptor in Digital Images”, Combinatorial Image Analysis. 14th International Worksho, IWCIA 2011.
- Roberto Guzmán, Oscar García-Olalla Olivera, Rocío Alaiz Rodríguez, ”Assessing and Visualizing the Stability of Feature Selectors: A Case Study with Spectral Data”, International Conference on Machine Learning and Data Analysis, Venice (Italy), 2011.
- Víctor González-Castro, Enrique Alegre, Sir Alexci Suárez-Castrillón, Oscar García-Olalla Olivera, María Teresa García-Ordás, ”Adaptative texture description for semen vitality assessment”, : Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas (International seminar of medical images processing and analysis) (SIPAIM), Bogotá (Colombia) 2010.
- Enrique Alegre, Maribel González, Víctor González-Castro, Tomás Alonso, María Teresa García-Ordás, Oscar García-Olalla Olivera, ”Evaluation of mother Wavelet functions with statistical texture descriptors to classify boar sperm acrosome”, Seminario Internacional de Procesamiento y Análisis de Imágenes Médicas (International seminar of medical images processing and analysis) (SIPAIM), Bogotá (Colombia) 2010.

Co-author in the following Intellectual properties and Patents

- Artificial vision procedure for the detection of proximal cytoplasmic droplets in spermatozoa[1], 2014. Procedimiento de visión artificial para la detección de gotas citoplasmáticas proximales en espermatozoides
- Artificial vision procedure for the detection of spermatozoa with curled tails[1], 2013. Procedimiento de visión artificial para la detección de espermatozoides con colas en ovillo.
- Detection of proximal droplets in tails of spermatozoa using artificial vision techniques[1], 2011. Reconocimiento de gotas proximales en colas de espermatozoide utilizando técnicas de visión artificial.
- Detection of distal droplets in tails of spermatozoa using artificial vision techniques[1], 2011. Reconocimiento de gotas distales en colas de espermatozoide utilizando técnicas de visión artificial.
- Classification of spermatozoa according to the state of their acrosomes using artificial vision techniques[1], 2011. Clasificación de espermatozoides en función del estado de su acrosoma utilizando técnicas de visión artificial.
- Detection of heads of spermatozoa with whip tails using artificial vision techniques[1], 2011. Reconocimiento de cabezas de espermatozoide con colas de látigo utilizando técnicas de visión artificial.

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.
  Financing entity: MICINN
  Duration from: 01/01/2013 to: 31/12/2015
  Main researcher: Enrique Alegre Gutiérrez
  Quantity: 84700 €
- ASASEC: Advisory System Against Sexual Exploitation of Children. European commission.
  Financing entity: UE
  Duration from: 16/10/2012 to: 16/10/2014
  Main researcher: Enrique Alegre Gutiérrez
  Quantity: 716.969,63 €
- Age estimation from face to detect people with special needs assistance. Telefónica.
  Financing entity: TELEFÓNICA

---

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

Duration from: 25/10/2013 to: 25/10/2014
Main researcher: Enrique Alegre Gutiérrez

- Automatic assessment of fresh and criopreserved boar sperm through digital image segmentation, analysis and classification. Spanish Ministry of Science and Innovation.
  Financing entity: MICINN
  Duration from: 01/01/2010 to: 31/12/2012
  Main researcher: Enrique Alegre Gutiérrez
  Quantity: 84000 €
- Proximal drops detection in boar spermatozoon tails through digital image processing. Microptic S.L.
  Financing entity: MICROPTIC S.L.
  Duration from: 01/10/2010 to: 31/08/2012
  Main researcher: Enrique Alegre Gutierrez
  Quantity: 10620 €
- Application of Leading Technologies to Unmanned Aircraft Research and Development in Air Traffic Management (ATLANTIDA)
  Main researcher: Jesús Gonzalo De Grado

Attended conferences

- XXXIV Jornadas de Automática, Terrassa, Spain, September 4-6 2013.
- International Conference on Similarity Search and Applications (SISAP), 2013.
- International Conference on Image Analysis and Recognition ICIAR. Aveiro, Portugal., 2012.
- Combinatorial Image Analysis. 14th International Worksho, IWCIA 2011
- III Eccomas thematic conference on computational vision and medical image processing. Algarve, Portugal, 2011

Summer schools

- Summer schools International Computer Vision Summer School (ICVSS 2012). 3R's of Computer Vision: Recognition, Registration, Reconstruction. Sicily, Italy. 16-21 July 2012
- International Computer Vision Summer School (ICVSS 2011). Registration, Recognition and Reconstruction in Images and Video Sicily, Italy. 11-16 July 2011

Annex B

SUMMARY OF THE THESIS IN SPANISH

RESUMEN DE LA TESIS EN CASTELLANO

En cumplimiento del punto 3° del artículo 19 del Reglamento de las enseñanzas oficiales de doctorado y del título de doctor de la Universidad de León, aprobado en Consejo de Gobierno el 25/9/2012, 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 análisis de texturas sigue siendo un problema abierto en el campo de la visión artificial. Cuando hablamos de análisis de textura nos referimos a un conjunto de algoritmos aplicados para detectar y describir variaciones espaciales del nivel de gris de los píxeles de una imagen. Hoy en día hay muchos campos que se dedican a la recuperación automática de texturas, de manera que los procesos se llevan a cabo de forma más rápida y sin necesidad de tanto personal cualificado. En esta tesis vamos a tratar tres aplicaciones diferentes en las que la información de la textura de las imágenes resulta muy útil: la evaluación de la calidad de los espermatozoides de verraco, la monitorización del desgaste de la herramienta de corte en procesos de fresado de borde y la recuperación de imágenes de textiles en escenas de un crimen o delito. Además, hemos estudiado el reconocimiento de rostros pero de manera muy breve.

1.1.1 Clasificación de espermatozoides de verraco

La evaluación de la calidad del semen en los procesos de inseminación artificial es indispensable en medicina y veterinaria. En esta tesis abordamos el problema de valorar la calidad del semen de verraco en la industria de la inseminación artificial. Los propietarios de granjas están muy interesados en el uso de muestras de espermatozoides de la mayor calidad posible para inseminar a las hembras. Las empresas de cría deben garantizar un nivel mínimo del potencial fertilizador de las hembras a la hora de comercializar las mismas. Con el objetivo de determinar la calidad de las muestras de semen, los expertos tienen en cuenta el número de espermatozoides muertos y también el porcentaje del total de cabezas que presenta el acrosoma dañado. Actualmente, la evaluación de estos parámetros, que determinan si una muestra es adecuada para inseminación artificial, se lleva a cabo usando un microscopio y tinciones para contar el número de espermatozoides válidos. Este es un proceso manual, muy caro debido al coste del microscopio de fluorescencia y también muy caro en cuanto a coste computacional. Además, es muy frecuente que se produzcan errores debido a la subjetividad del experto que analiza las muestras. Añadido a todo esto, nos encontramos con que, debido al proceso de tinción, el número de espermatozoides muertos aumenta, de manera que los resultados ya no son tan fiables. Por todo esto, se hace necesario el desarrollo de un proceso automático para detectar acrosomas dañados y espermatozoides muertos que no dependa de la utilización de tinciones. Este método se llevó a cabo mediante el procesamiento de imágenes digitales en escala de grises extrayendo información relevante por medio del análisis de texturas. De esta manera el proceso es más rápido, más fiable y además permite que los criaderos ahorren dinero, ya que los requisitos son simplemente una cámara digital, un ordenador y un microscopio convencional.

1.1.2 Monitorización del desgaste de la herramienta de corte en procesos de fresado

El desarrollo de técnicas de control para la detección del desgaste de herramientas de corte en procesos de mecanizado es de gran utilidad para los sistemas de producción. La monitorización del estado de las herramientas ayuda a reducir costes drásticamente, bien porque se reduce el tiempo en que un operador tiene que analizar la herramienta o bien ayudando a prevenir roturas que el operador puede haber pasado por alto. En los últimos años, la búsqueda de técnicas que estimen el desgaste de las herramientas, se ha convertido en un factor clave debido, sobre todo, a la aparición de máquinas de corte muy rápidas, que hacen que la vida de la herramienta se vea reducida. La Universidad de León participó en el proyecto de investigación FRESVIDA, fundado por el Ministerio de Economía y Competitividad con el objetivo de encontrar un sistema que determine la vida de la herramienta en procesos de fresado usando simplemente la información obtenida por imágenes de la región de desgaste. En este contexto, el uso de la visión artificial y, en concreto, el análisis de texturas, es muy interesante para desarrollar un sistema que sea preciso y eficiente y que sea capaz de determinar el estado de la herramienta de corte en tiempo real. Además, el sistema no se verá afectado por problemas como los que afectan a los métodos indirectos, como pueden ser el ruido que influye en los sensores de fuerza, las vibraciones, etc.

1.1.3 Recuperación de imágenes basadas en su contenido para la recuperación de textiles en habitaciones

La recuperación de textiles constituye una herramienta clave en el ámbito de la pornografía infantil. La búsqueda de textiles que estén presentes en alguna escena de grandes conjuntos de imágenes formados por casos de pornografía infantil, permite que se puedan recuperar imágenes y vídeos que probablemente estarán relacionados con el caso de interés. Normalmente, los criminales ocultan su rostro para evitar que se les reconozca pero no tienen en cuenta el resto de información útil que se puede extraer de cada imagen. El reconocimiento de textiles tiene una gran ventaja, y es que rara vez estos textiles son movidos o cambiados de lugar en las escenas. Por eso, la recuperación de textiles es tan útil.

1.2 Objetivos

El principal objetivo de esta tesis es la evaluación de algoritmos de descripción de textura ya existentes y el desarrollo de nuevos descriptores que mejoren a los métodos del estado del arte en diversas aplicaciones. Dado este objetivo general, se pueden definir los siguientes objetivos particulares:

1. Mejorar las tasas de acierto de los descriptores de textura conocidos en experimentos de reconocimiento de materiales utilizando conjuntos de imágenes que supongan un reto, como KTH TIPS2-a o Brodatz.

2. Incrementar la efectividad en las tareas de clasificación de espermatozoides, como por ejemplo, en la clasificación en función de la vitalidad y en la identificación de acrosomas dañados, utilizando técnicas de descripción de textura.

3. Evaluar automáticamente del desgaste de herramientas de corte en procesos de fresado teniendo únicamente en cuenta información de su superficie.

4. Recuperar imágenes basándonos en su contenido, en este caso, recuperar textiles por medio de un método completamente automático dentro del marco del proyecto europeo ASASEC.

1.3 Contribuciones principales

Las principales contribuciones de esta tesis se pueden resumir como sigue:

1. Se ha propuesto un nuevo método llamado ALBPS (Adaptive Local Binary Patterns with oriented Standard deviation) basado en el método presentado como Guo, Zhang, Zhang and Zhang (2010). Fue evaluado para el reconocimiento de materiales y para analizar la vitalidad de espermatozoides y se obtuvieron resultados satisfactorios.

2. Se ha desarrollado un nuevo potenciador para la descripción de texturas llamado LOSIB (Local Oriented Statistical Information Booster). La evaluación del mismo se llevó a cabo con los conjuntos de imágenes de reconocimiento de materiales que suponían un mayor reto así como con algunos descriptores de la literatura.

3. Para evaluar el estado del acrosoma de espermatozoides se propuso una fusión temprana de LOSIB con características de Fourier y Haralick extraídas de las respuestas Wavelet. Este método mejora los trabajos anteriores y también los resultados obtenidos por cada descriptor por separado.

4. Se presentó una versión más compleja de LOSIB llamada CLOSIB (Complete LOSIB). CLOSIB añade nuevas funcionalidades. Por ejemplo, permite la inclusión de información sobre la media o la desviación estándar orientada de los valores de gris de una imagen y de varios vecindarios. Se propusieron tres variantes de CLOSIB: H-CLOSIB, M-CLOSIB y HM-CLOSIB. Las principales diferencias entre estos métodos y CLOSIB son el número de orientaciones que se tienen en cuenta (Half CLOSIB) y la disposición de los vecinos para diferentes radios (Multiscale CLOSIB).

5. Se ha presentado un nuevo método CBIR (Content Based Image retrieval) para el reconocimiento de tejidos como alfombras, cortinas, etc. Este método extrae la descripción local de la textura de las regiones máximas estables (MSER) de las imágenes de la escena.

1.4 Organización de la tesis

En esta sección se describe la estructura de la tesis doctoral. El primer capítulo introductorio se ha centrado en la motivación de esta disertación, los objetivos principales y las contribuciones más relevantes. El resto de los capítulos están organizados como sigue: El capítulo 2 presenta una revisión de los principales métodos del estado del arte relacionados con la descripción basada en texturas. Nos centramos en los trabajos referentes a aplicaciones que tienen relación con esta tesis. Por esta razón, hemos evaluado métodos recientes para el reconocimiento de materiales publicados en las revistas y conferencias más relevantes. También hemos estudiado la literatura relacionada con la descripción y clasificación de espermatozoides utilizando técnicas de visión artificial. Además, se comentan algunos de los métodos publicados que tratan con el problema de la monitorización del desgaste de las herramientas de corte. En este capítulo se introduce también una pequeña evaluación de trabajos relacionados con el reconocimiento de rostros. Finalmente, se revisan métodos de descripción de texturas para reconocimiento de tejidos, lo cual demuestra lo novedoso de la aplicación.

En el capítulo 3 se describen brevemente unos de los métodos de descripción de imágenes más populares en el estado del arte de las últimas décadas. Hemos explicado el método LBP (Local Binary Patterns) y algunas de sus modificaciones más populares: ALBP, CLBP y LBPV. Además se introducen algunos descriptores globales basados en la transformada Wavelet y descripción de forma.

Los conjuntos de imágenes empleados en los experimentos se muestran en el capítulo 4. Hemos descrito dos conjuntos de imágenes para reconocimiento de materiales: KTH TIPS2-a and Brodatz32. También explicamos la creación de dos conjuntos de imágenes para la evaluación de la vitalidad o del estado del acrosoma de espermatozoides. Después, describimos tres conjuntos de imágenes usados para monitorización del desgaste de la herramienta y el conjunto JAFFE para el reconocimiento de rostros. Finalmente, se presenta un conjunto desarrollado para la recuperación de textiles utilizando imágenes extraídas de vídeos disponibles públicamente en Youtube, llamado TextilTube.

El capítulo 5 muestra nuestra primera contribución para la descripción de texturas llamada ALBPS (Adaptive Local Binary Patterns). Introduce una modificación al método ALBP, ya que incluye información complementaria sobre el gradiente en diferentes orientaciones. En este capítulo empleamos el nuevo método en dos aplicaciones: reconocimiento de materiales y clasificación de los espermatozoides según su vitalidad.

En el capítulo 6 se introduce el algoritmo LOSIB (Local Oriented Statistical Information Booster). Este método se usa en combinación con descriptores de texturas incrementando su eficiencia en numerosos experimentos relacionados con el reconocimiento de materiales, la integridad de los acrosomas y la monitorización del desgaste de las herramientas.

Nuestra versión extendida de LOSIB, llamada CLOSIB (LOSIB Completo), se describe con detalle en el capítulo 7. Además, se introducen y se comparan tres variantes de CLOSIB en la sección de experimentos: Half CLOSIB, Multi Scale CLOSIB y Half Multi Scale CLOSIB. Este método ha sido evaluado en diferentes aplicaciones, como reconocimiento de materiales, de rostros y también de textiles.

El capítulo 8 contiene un resumen con las conclusiones de esta tesis y da una visión sobre las posibles líneas de trabajo futuro relacionadas con ella.

2 Revisión del estado del arte

La clasificación de texturas sigue siendo un problema abierto en el ámbito de la visión por computador y reconocimiento de patrones. Muchas compañías están intentando aumentar la productividad integrando procesos automáticos gestionados por cámaras. Además, la policía y otras fuerzas de seguridad tienen que tratar cada día con miles de imágenes en las que tienen que detectar un posible crimen, algún objeto específico,... Una de las mejores estrategias para detectar un objeto en una imagen es mediante la información de textura, como se ha demostrado en un buen número de trabajos (Liang and Juang, 2015),(Nanni et al., 2012),(Dornaika et al., 2014). Además, la descripción basada en textura también es muy útil para procesos de control de calidad, detección de peatones en calles abarrotadas (Castrillon-Santana et al., 2015) o incluso ayudando a los médicos con las interpretaciones de las imágenes médicas (Hegenbart et al., 2013),(Kavitha et al., 2015).

2.1 Descripción de texturas

Las propiedades como el color, la forma y la textura son fundamentales para la descripción de los objetos. Hoy en día, los métodos de descripción de textura morfológica, se emplean mucho para imágenes en color o en escala de grises. En la revisión presentada en (Aptoula and Lefevre, 2011), se revisan una gran cantidad de técnicas de descripción de textura. En (Drbohlav and Leonardis, 2010), las texturas se describen como estructuras básicas adyacentes, cuya colocación y orientación se rigen por ciertas reglas de generación. Junto con el color y la forma, la textura constituye una de las tres propiedades fundamentales de los objetos y por esta razón ha sido un objeto de estudio para la comunidad de análisis de imágenes desde sus primeros días. Este tipo de análisis ha encontrado aplicaciones en muchos campos, incluyendo pero no limitado a la anotación y recuperación de imágenes basadas en contenido (CBIR), medicina y ciencias naturales.

En esta tesis, nosotros nos centramos en la descripción de textura, cuyo principal objetivo es caracterizar una textura dada, normalmente con el objetivo de asignarla a una determinada clase.

Según la revisión realizada por (Aptoula and Lefevre, 2011), podemos dividir los enfoques morfológicos de la descripción de la textura en cinco categorías: series morfológicas y características de distribución de tamaños, distribuciones de orientación-distancia, distribución invariante a rotación e iluminación, distribuciones de forma-tamaño y de distancia-orientación-tamaño.

2.1.1 Ventajas de las características de la textura morfológica

En análisis morfológico es muy útil para texturas. Está basado en la relación espacial entre píxels y cuenta con muchas herramientas para extraer información de tamaño y forma. Esto es espacialmente valorado cuando nos centramos en las primitivas de la textura. Por otra parte, las herramientas morfológicas multi escala son relativamente fáciles de implementar, gestionando así las variaciones primitivas del tamaño de la textura, mientras que los operadores morfológicos conectados representan un potente conjunto adicional de herramientas capaces de explotar las conectividades de los píxeles.

2.1.2 Series morfológicas y características de distribución de tamaños

La mayoría de las características morfológicas de la textura se basan en el principio de series morfológicas, que conducen a distribuciones unidimensionales o multidimensionales, basadas en una o más propiedades de SE, tales como tamaño, forma, orientación, etc. La aplicación de un filtro morfológico tal como una abertura estructural con una SE de tamaño creciente λ se resuelve en una serie de imágenes filtradas sucesivas con cada vez menos detalles. Los momentos estadísticos de orden superior pueden ser calculados para formar el vector característico final, algunos de los cuales son habituales, incluyendo la media, la varianza, la asimetría y la curtosis.

2.1.3 Distribuciones de distancia-orientación

Aunque la granulometría y el espectro de patrones dependen de operaciones de aperturas y cerramientos, también es posible involucrar a cualquier operador morfológico, como la erosión, para construir una característica de textura morfológica capaz de describir el contenido de textura global. Un claro ejemplo en este sentido es la característica de covarianza morfológica K, definida como el homólogo morfológico del operador de autocorrelación. En este caso, el SE b en consideración, consiste en un conjunto de dos puntos p_1 y p_2 y ambos definen un tamaño $2\lambda = |p_1 p_2|$ y una orientación $\vec{v} = p_1\vec{p}_2/|p_1\vec{p}_2|$:

$$K^{\vec{v}}(f) = K^{\vec{v}}_{\lambda}(f)|K^{\vec{v}}_{\lambda}(f) = \Sigma_{p\in E}\Pi^{\xi}_{\lambda,\vec{v}}(f)(p) \tag{1}$$

donde

$$\xi'_{\lambda,\vec{v}}(f)(p) = f(p-\lambda\vec{v}) \bigwedge f(p+\lambda\vec{v}) \tag{2}$$

2.1.4 Distribuciones invariantes a iluminaciones e rotaciones

Aunque el uso de varias orientaciones de pares de puntos transforma la covarianza morfológica en una valiosa herramienta anisotrópica, en la práctica generalmente se prefiere describir una textura sin tener en cuenta su orientación. En Aptoula and Lefevre (2011), van un paso más allá con un conjunto de nuevos retos, de los cuales la descripción de la textura completamente invariante es un elemento importante (Shotton et al., 2009), y por eso en la última década se ha notado un aumento de la demanda de todos los tipos de métodos invariantes en el contexto del análisis de textura, con respecto a la escala, el punto de vista y las condiciones de iluminación. Por otro lado, aunque las series morfológicas son adecuadas para el análisis de escalas múltiples, y los espacios de la escala morfológica han sido estudiados a fondo, la rotación y especialmente la descripción de la iluminación invariante de la textura morfológica por otro lado no ha recibido la atención que merece.

2.1.5 Distribuciones tamaño-forma

En este tipo de distribuciones, se considera un único parámetro λ para medir la evolución de tamaño a través del SE $b\lambda$. Esta definición, asumiendo un único parámetro que varía de tamaño λ, nos impide realizar mediciones precisas. De hecho, no es adecuado para granos de textura elíptica o rectangular, por ejemplo, donde se deben tener en cuenta los dos ejes independientes. Así, se han realizado varios intentos para construir series morfológicas bi-variadas, lo que permite obtener mediciones de tamaño y forma. De esta manera, citeLefevre07 consideran elementos estructurantes con dos parámetros de tamaño diferentes α y β que varían de forma independiente. En Ghosh and Chanda (1998) se propuso un enfoque similar. Introducen operadores morfológicos paramétricos condicionales y construyen un conjunto 2D de SE con un tamaño creciente, tanto en las dimensiones horizontal como vertical.

2.1.6 Distancia-Orientación-Tamaño

La covarianza extrae un vector de características que contiene información sobre periodicidad y direccionalidad, mientras que la granulometría se concentra más bien en la granularidad de su entrada. En consecuencia, ambos son necesarios en el caso general para una descripción eficaz de la textura. Sin embargo, su combinación es bastante ambigua, ya que se puede realizar de varias maneras. El método más evidente o directo, es calcular independientemente cada vector de características y luego concatenar ambos. Se ha propuesto una alternativa en Aptoula and Lefevre (2007), que consiste en unificar las funcionalidades de los dos operadores al variar en paralelo tres propiedades SE: su tamaño, dirección y distancia.

2.1.7 Métodos más comúnmente usados

Las características de Haralick han sido ampliamente usadas durante los últimos 30 años para la descripción de texturas. Recientemente, Chaddad et al. (2011) desarrollaron un sistema que emplea estas características para detectar células de cáncer de colon. De forma similar, Ribaric and Lopar (2012), llevaron a cabo una extracción de características locales de Haralick en una aplicación de impresión de la palma con resultados satisfactorios. Además, también se han desarrollado métodos basados en Wavelet en los últimos años, demostrando su eficiencia en problemas de recuperación de texturas. En Carbunaru et al. (2009), los autores proponen un sistema para la recuperación de imágenes de textiles usando ICA (Independent Component Analysis) aplicado a la respuesta de la transformada Wavelet y obteniendo una precisión del $89\,\%$-$94\,\%$. Últimamente, los descriptores locales han ido creciendo en importancia. Más concretamente, el LBP (Local Binary Patterns), propuesto por Ojala et al. (1994) ha sido muy empleado debido a su simplicidad y su gran capacidad para extraer las características intrínsecas de las texturas. Guo y su grupo de investigación llevaron a cabo algunas modificaciones de LBP como LBP Varianza (LBPV) (Guo, Zhang and Zhang, 2010), LBP Completo (CLBP)(Guo, Zhang and Zang, 2010) o LBP Adaptativo (ALBP)(Guo, Zhang, Zhang and Zhang, 2010).

2.2 Campos de aplicación

A continuación vamos a estudiar los diferentes campos de aplicación donde nuestras propuestas han demostrado mejoría: clasificación de espermatozoides, monitorización de desgaste de herramientas, reconocimiento de materiales, reconocimiento facial y recuperación textil.

2.2.1 Clasificación de espermatozoides

La fertilidad del rebaño es fundamental para la prosperidad de cualquier criadero de animales. La inseminación artificial es a la vez un tratamiento de fertilidad para los seres humanos y un método típico en la cría de cerdos o ganado lechero, que domina el proceso reproductivo de muchas granjas.

Hasta donde conocemos, la única manera de medir automáticamente el estado y la vitalidad de una muestra de esperma es el uso de imágenes de fluorescencia en combinación con los contadores de células o microscopios de fluorescencia, que es un dispositivo muy caro. De lo contrario, las muestras teñidas deben evaluarse manualmente, lo que requiere la presencia de expertos veterinarios y el uso equipos especializados que lleven a una tarea costosa y no objetiva. Este trabajo trata de la evaluación automática y confiable del estado del acrosoma de espermatozoides de verraco usando imágenes de contraste de fase, por lo que sólo se necesita un microscopio de contraste de fase, una cámara digital de alta gama y un ordenador.

Recientemente, en 2012, Alegre et al. (2012) alcanzó una tasa de éxito del 93,2 %. Alegre et al. (2013) mejoró este enfoque utilizando 7 contornos interiores y calculando un descriptor de textura local para cada punto de los siete contornos y clasificándolo con LVQ, obteniendo el mejor resultado hasta el momento con una tasa de éxito del 99 %. No sólo la descripción de las células es importante, sino que también la segmentación de las cabezas, si las hay, y la fase de clasificación pueden ser críticas. En este sentido, Gonzalez-Castro et al. (2009), propusieron un método de segmentación novedoso e inteligente basado en un umbral cambiante y en una cuenca hidrográfica con la que el 90,96 % de 763 imágenes de espermatozoides se consiguieron segmentar correctamente. Bijar et al. (Bijar et al., 2012), segmentaron el acrosoma de los espermatozoides humanos mediante un método basado en un clasificador bayesiano que utiliza el método de mezclas adaptativas (AMM) y el modelo de campo aleatorio de Markov (MRF) para obtener y actualizar la densidad de probabilidad condicional de la clase (CCPDF) y la probabilidad a priori de cada clase. Por lo que respecta a la clasificación, los métodos de clasificación no supervisados demostraron tener un mejor rendimiento que los supervisados (Alegre, Gonzalez-Castro, Suarez and Castejon-Limas, 2009). Sin embargo, además de estos experimentos, es muy importante obtener una clasificación precisa. En González-Castro et al. (2012) y Alegre, González-Castro, Suárez and Castejón (2009) se estima la proporción verdadera y desconocida de células dañadas en una muestra con ayuda de la distancia de Hellinger cuantificando las probabilidades a priori desconocidas de los conjuntos de prueba.

2.2.2 Monitorización del desgaste de la herramienta

El desarrollo de sistemas de medición en línea para detectar el nivel de desgaste de las plaquitas de corte es una cuestión de suma importancia para el control de sistemas de producción automatizados. Este tipo de sistemas de monitorización ayudan a reducir los costes de producción, ya que ya no se requiere la supervisión de un operador. Este tema se hace cada vez más importante en instalaciones modernas donde el mecanizado se realiza a velocidades de corte muy altas, lo que reduce significativamente la vida de la herramienta. Como consecuencia, el coste de reemplazo de herramientas reduce la productividad. A pesar de los muchos esfuerzos que se han centrado en este tema, no se ha logrado encontrar una solución satisfactoria para la monitorización en línea debido a las grandes dificultades que implica la medición del desgaste de las herramientas. Las operaciones de mecanizado de metales como torneado, fresado o perforación son procesos involucrados en la fabricación de la mayoría de las mercancías. Aunque muchas investigaciones se han centrado en adquirir conocimientos avanzados sobre el comportamiento de estos procesos, en las últimas décadas, la complejidad de este trabajo en este campo hace que este desafío siga vivo hoy en día (Bhuiyan and Choudhury, 2014).

En particular, en los últimos años se ha centrado la investigación en el uso de diferentes sensores para monitorizar el estado de las herramientas de corte. Algunas de los trabajos más recientes son (Painuli et al., 2014; Wang et al., 2014; Pratama et al., 2011; Fernández-Abia et al., 2014; Fayad, 2010; Gao et al., 2010; Li et al., 2010; Qian et al., 2010). El uso de sensores de desgaste puede reducir los costes de la herramienta en un 40 % (Lim, 1995), lo cual representa una clara oportunidad de negocio.

Por lo tanto, es muy importante desarrollar métodos precisos y eficientes para predecir el nivel de desgaste de la herramienta, su evolución y su vida de corte restante. Un sistema de control debe ser capaz de proporcionar estrategias optimizadas para el reemplazo de herramientas y para ajustar correctores de herramientas. En este campo, el uso de la visión por ordenador y la análisis de la textura es de interés para determinar el desgaste de las plaquitas de corte.

2.2.3 Reconocimiento de materiales

KTH Tips2-a es un conjunto de datos que se ha vuelto muy popular en los últimos años para el reconocimiento de materiales. Fue creado por Caputo et al y presentado en (Caputo et al., 2005). Es uno de los conjuntos de datos de reconocimiento de textura más difíciles debido a la gran distancia intraclase de cada material.

Por esta razón, varios investigadores están probando sus métodos con él, como por ejemplo, Chen, Tian, Lee, Zheng, Smith and Laine (2010), quienes propusieron un método llamado MWLD basado en la Ley de Weber que logró Un 64,7 % de tasa de acierto en este conjunto de datos. Así mismo, Sharma et al. (2012) presentó un método denominado LHS (Local Higher-Order Statistics), que obtiene una precisión del 73,0 %. En esta misma línea, tenemos el método AMBP propuesto por Hafiane et al. (2015) basado en LBP que logró un 70,3 % de precisión. Como podemos ver, todavía hay margen de mejora en relación con este conjunto de datos específicos. Otro conjunto de datos ampliamente utilizado es el conjunto de datos Brodatz32 (Valkealahti and Oja, 1998), que contiene imágenes en escala de grises de 32 texturas con ataques de rotación y escalado. Hemos probado nuestra propuesta con estos dos conjuntos de datos, ya que consideramos que son muy representativos para problemas de textura.

2.2.4 Reconocimiento de rostros

El reconocimiento facial es otro campo interesante en el que la descripción de la textura ha demostrado ser muy útil. Miles de imágenes que contienen caras se comparten a través de internet todos los días. La posibilidad de reconocer su identidad es muy valiosa para múltiples aplicaciones. En Elaiwat et al. (2015), se propuso una técnica basada en la transformada de Curvelet para el reconocimiento facial 3D. Marsico et al. (2015) desarrolló un método que utiliza la descripción de regiones para reconocer caras después de una cirugía plástica con resultados muy interesantes. En Shenoy et al. (2015), los autores propusieron una nueva técnica para el reconocimiento facial basado en una fusión de características Wavelet y Fourier. El procedimiento de evaluación se llevó a cabo utilizando varias bases de datos faciales, incluyendo JAFFE. En la misma línea de trabajo, Wan et al. (2015) logró un 79 % de tasa de éxito utilizando un nuevo método denominado Q-SVD + RWN.

2.2.5 Recuperación de imágenes basada en su contenido para textiles

El principal interés de los métodos CBIR consiste en recuperar imágenes de una colección que coinciden con la preferencia de un usuario. Las imágenes se pueden describir utilizando características, tales como color, forma, textura o cualquier otro tipo de información que se puede extraer automáticamente de ellos. Se han obtenido resultados interesante utilizando este tipo de técnicas en varios campos de investigación como imagen médica (Faria et al., 2015; Srinivas et al., 2015; Bugatti et al., 2014), biológicos (Feng et al., 2016; Mallik et al., 2010) o seguridad biométrica (Iqbal et al., 2012).

Varios grupos de investigación están tratando de hacer frente al reconocimiento de objetos utilizando diferentes enfoques como ”template matching” (Shih and Yu, 2015; Tan et al., 2016), color (Li et al., 2015; Zhu et al., 2015), o características locales invariantes (SIFT, SURF, etc.) (Chang et al., 2012; Chen et al., 2015). Sin embargo, la mayoría de estas técnicas fallan cuando el objeto es demasiado plano o deformable, cuando hay modificaciones de color debido a cambios de iluminación o cuando no hay suficientes puntos clave para describirlo. Estos problemas suelen aparecer cuando se trata de textiles en lugar de recuperación de objetos. Los mismos textiles pueden aparecer en las imágenes en múltiples condiciones y con múltiples formas diferentes. Por esta razón, los sistemas clásicos de reconocimiento de objetos no son los más adecuados para recuperar este tipo de objetos.

Recientemente, en Chun and Kim (2013), se propone un nuevo método que, en contraste con los otros trabajos descritos anteriormente, realiza la recuperación utilizando un gran conjunto de datos compuesto por 1343 imágenes textiles. En Huang and Lin (2014) se propuso un sistema basado en la combinación de características de color, textura y forma para recuperar textiles sobre más de 4000 imágenes descargadas de Globle-Tex Co. El sistema de recuperación se basó en un proceso de firma extraído usando el método de agrupamiento k-means y obteniendo un 83 % de tasa de acierto. Todos los sistemas de recuperación de materiales funcionan bien con fines de construcción, pero no lograrán un buen resultado en nuestro problema de recuperación de textiles. Las dos diferencias principales son, en primer lugar, el tamaño de las regiones en la imagen. En nuestro caso, los textiles pueden estar formados por unos pocos píxeles. Y en segundo lugar, la heterogeneidad de los textiles en comparación con los materiales de construcción que son típicamente muy similares a lo largo de toda su superficie.

Varios grupos de investigación están tratando de encontrar el mejor algoritmo para la segmentación de la región. En 2016, Zheng y Sarem propusieron un método llamado NAMES que se basa en la idea de agrupar píxeles con un rendimiento muy alto en términos de tiempo (Zheng and Sarem, 2016). En Yang et al. (2015) se propuso un método basado en la segmentación del histograma de color utilizando el espacio de color HSV. En Matas et al. (2004) propusieron un método basado en la extracción de regiones estables máximas teniendo en cuenta un umbral binario que varía a lo largo de todo el espectro de escala de grises.

3 Metodología

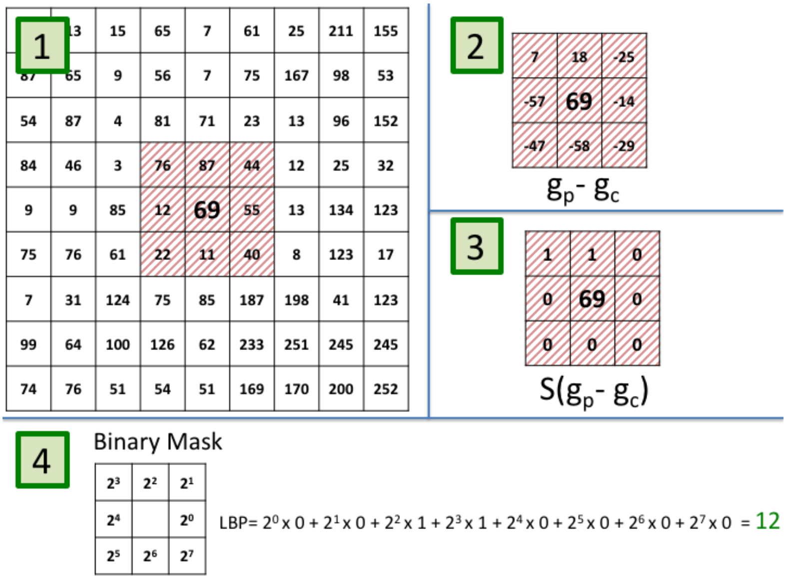
En este capítulo vamos a hacer un breve resumen de los métodos que se han utilizado a lo largo de esta tesis junto con los métodos propuestos.

3.1 Descriptores locales

Local Binary Patterns (LBP) (Kashyap and Khotanzed, 1986) es un operador de textura simple y eficiente que etiqueta cada píxel de la imagen analizando su vecindario, estudiando si el nivel de gris de cada píxel supera un determinado umbral y codificando dicha comparación mediante un número binario. Debido a su bajo coste computacional y al gran poder discriminativo que ha demostrado tener, LBP se ha convertido en los últimos años en una de las soluciones más utilizadas en numerosas aplicaciones relacionadas con textura. La característica más importante de LBP en aplicaciones reales es la robustez que ofrece frente a variaciones en la intensidad del nivel de gris, causado, entre otras muchas cosas, por diferencias en la iluminación. Para calcular LBP sobre una imagen en escala de grises se utiliza la ecuación 3.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \ , \ s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} \tag{3}$$

donde P es el número de vecinos que se van a considerar, R es el tamaño del vecindario y, g_c y g_p son los valores de gris del píxel central y cada uno de los p píxeles del vecindario respectivamente. En la figura 1 podemos observar un ejemplo del cálculo del LBP de manera gráfica sobre un pixel cualquiera de una imagen para los parámetros P=8 y R=1.



	13	15	65	7	61	25	211	155
87	65	9	56	7	75	167	98	53
54	87	4	81	71	23	13	96	152
84	46	3	76	87	44	12	25	32
9	9	85	12	69	55	13	134	123
75	76	61	22	11	40	8	123	17
7	31	124	75	85	187	198	41	123
99	64	100	126	62	233	251	245	245
74	76	51	54	51	169	170	200	252

7	18	-25
-57	69	-14
-47	-58	-29

1	1	0
0	69	0
0	0	0

2^3	2^2	2^1
2^4		2^0
2^5	2^6	2^7

$LBP = 2^0 \times 0 + 2^1 \times 0 + 2^2 \times 1 + 2^3 \times 1 + 2^4 \times 0 + 2^5 \times 0 + 2^6 \times 0 + 2^7 \times 0 = 12$

Figura 1: Esquema de los pasos necesarios para obtener el valor de LBP de un píxel concreto.

En 2002, se realizó una nueva propuesta por parte de varios de los autores del método original (Ojala et al., 2002), al darse cuenta de que la mayoría de la información relevante de la textura era descrita mediante un patrón uniforme. Se puede definir patrón uniforme como aquel patrón que tiene como mucho dos saltos entre 0 y 1 o entre 1 y 0 en su cadena, asumiéndose que es un patrón circular, por lo que es necesario comprobar también si hay transición entre el último elemento y el primero. Teniendo en cuenta únicamente estos patrones uniformes, cualquier píxel de la imagen se puede representar mediante P+2 valores diferentes, asignando desde 0 hasta P el valor que corresponde con el número de píxel cuyo valor es 1, y el valor P+1 a cualquier patrón no uniforme. En la ecuación 4 se puede ver de manera matemática el cálculo del LBP uniforme.

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c) & \text{if} \quad U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \tag{4}$$

donde

$$U(LBP_{P,R}) = |s(g_{P-1} - g_c) - s(g_0 - g_c)| + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \tag{5}$$

En los últimos años han surgido bastantes métodos basados en LBP. Un ejemplo es ALBP de su nombre en inglés *Adaptive Local Binary Patterns* (Guo, Zhang, Zhang and Zhang, 2010), centrado en incorporar información acerca de la orientación de las texturas. El método trata de minimizar las variaciones de la media y la desviación estándar de las diferencias direccionales (diferencias del valor de gris del píxel central y cada uno de los vecinos dependiendo de su posición respecto al centro) . Para ello, se propuso añadir un parámetro extra w en la ecuación $|g_c - w_p * g_p|$. La función objetivo es definida según la ecuación 6.

$$w_p = arg_w \min \left\{ \sum_{i=1}^{N} \sum_{j=1}^{M} |g_c(i,j) - w \cdot g_p(i,j)|^2 \right\}, \tag{6}$$

donde w_p es el elemento utilizado para minimizar la diferencia de la dirección p y, N y M son el número de filas y de columnas de la imagen. Una vez dicho esto, la ecuación final del Adaptive Local binary Patterns (ALBP) se define como:

$$ALBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - w_p \cdot g_c)2^p \ , \ s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases} . \tag{7}$$

.

Otro método, Local Binary Patterns Variance fue propuesto por Guo y su grupo de investigación como una combinación de LBP y un método de distribución de contraste (Guo, Zhang and Zhang, 2010). Este método utiliza la varianza de la imagen como peso adaptativo para ajustar la contribución de cada valor de LBP en el cálculo del histograma. LBPV se calcula de la siguiente manera:

$$LBPV_{P,R}(k) = \sum_{i=1}^{N} \sum_{j=1}^{M} w(LBP_{P,R}(i,j),k), k \in [0,K] \tag{8}$$

Donde k es cada uno de los valores del histograma, K es el valor máximo de LBP y w es definido como:

$$w(LBP_{P,R}(i,j),k) = \begin{cases} VAR_{P,R}(i,j), & LBP_{P,R}(i,j) = k \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

donde $VAR_{P,R}$ es la varianza del vecindario:

$$VAR_{P,R} = \frac{1}{P} \sum_{p=0}^{P-1} (g_p - u)^2 \tag{10}$$

siendo μ la media del vecindario.

El mismo grupo de investigación que desarrolló LBPV y ALBP, presentó otro método llamado CLBP (Compound Local Binary Patterns) que trata de generalizar y completar la información que aporta el LBP clásico (Guo, Zhang and Zang, 2010). En este método, una región local de la imagen es representada por su píxel central y una transformada local de diferencias signo-magnitud llamada LDSMT por sus iniciales en inglés. LDSMT descompone estructura local de la imagen en dos componentes complementarios: la diferencia de signos, que se corresponde con el clásico LBP y lo llaman CLBP_S y la diferencia de magnitudes CLBP_M que se define utilizando la ecuación 11.

$$CLBP\_M_{P,R} = \sum_{p=0}^{P-1} t(m_p,c)2^p \ , \ t(x,c) = \begin{cases} 1 & \text{if } x \geq c \\ 0 & \text{if } x < c \end{cases} \tag{11}$$

siendo c un umbral determinado adaptativamente. Normalmente se utiliza el valor medio de todas las diferencias de magnitudes de la imagen.

HOG (Histogram of Oriented Gradients) (Dalal and Triggs, 2005) evalúa histogramas locales de las orientaciones del gradiente de una imagen sobre una rejilla. El concepto principal de este método esta basado en la idea de que la apariencia local de caras, peatones, o cualquier tipo de objeto a menudo se describen teniendo en cuenta la distribución local de los bordes. El método está implementado dividien-

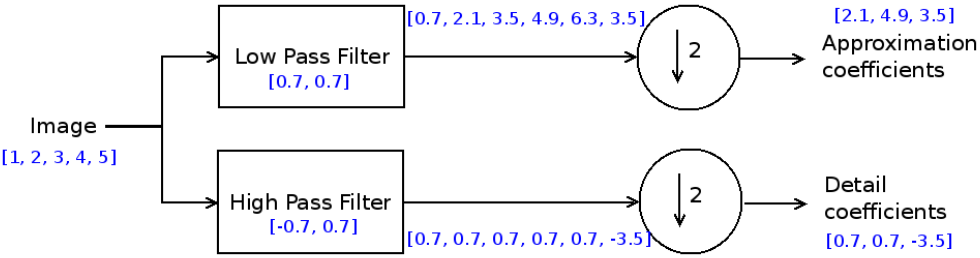


Figura 2: Diagrama representando el primer nivel de descomposición de DWT

do una imagen en pequeñas regiones uniformes llamadas celdas que suelen estar solapadas entre sí. A continuación, para cada celda, se calcula el histograma de las orientaciones del gradiente. El descriptor final está formado por la concatenación de todos los histogramas de las celdas de la imagen.

3.2 Descriptores globales

En esta tesis hemos utilizado también la Transformada Discreta de Wavelet (DWT) debido al gran rendimiento que ha demostrado tener a la hora de describir texturas. De una manera sencilla, la transformada Wavelet descompone una imagen en la suma de una imagen general y una imagen detallada. El DWT de una imagen I se calcula aplicando filtros de paso bajo y paso alto g como en la ecuación 12.

$$y[n] = (I * g)[n] = \sum_{-\infty}^{\infty} x[k]g[n-k] \tag{12}$$

Cuando a la imagen se le pasa un filtro de paso bajo, solo la información general se mantiene. Por el contrario si la imagen es convolucionada con un filtro de paso alto, los pequeños detalles de la imagen se mantienen. A esto se le conoce como nivel 1 de descomposición. A cada una de estas imágenes además, se le puede aplicar de nuevo el filtrado, creando nuevos niveles de descomposición. Se puede ver un diagrama con una dimensión en la figura 2.

En nuestro caso hemos elegido los Wavelets de Haar para calcular la DWT. Los wavelet de Haar introducidos en Haar (1910), es el wavelet mas simple posible. El conjunto de funciones puede ser descrito por su función madre $\psi(t)$ como:

$$\psi(t) = \begin{cases} 1 & \text{if } 0 \leqslant t < 1/2 \\ -1 & \text{if } 1/2 \leqslant t < 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

con función de escalado $\phi(t)$:

$$\phi\left(t\right)=\begin{cases}1 & \text{if } 0\leqslant t<1\\ 0 & \text{otherwise}\end{cases} \tag{14}$$

Para obtener el descriptor de textura, se calcularon diferentes métricas a partir de la matriz de concurrencia del nivel de gris (GLCM) de la imagen original y de las cuatro sub imagenes del primer nivel de descomposicion del DWT de Haar, como hicieron Arivazhagan and Ganesan (2003).

A partir de la matriz GLCM se pueden obtener múltiples descriptores. En Haralick (1979) se propusieron 14 características de las cuales hemos cogido 13 dejando fuera el Coeficiente de correlación máximo. Estas 13 características se han calculado para cada una de las 5 imágenes obteniéndose un vector de 65 características llamado WCF13 o WCF4 en el caso de escoger solo 4 de ellas (energía, contraste, correlación y momento diferencial inverso).

4 Conjunto de imágenes

4.1 Conjuntos de imágenes de recuperación de materiales

Para la recuperación de materiales se han utilizado dos conjuntos de imágenes diferentes: KTH TIPS2-a y Brodatz32.

KTH TIPS2-a está compuesto por 4752 imágenes específicas para reconocimiento de materiales (Caputo et al., 2005). Contiene 11 materiales diferentes (lechuga, pan marrón, pan blanco, aluminio, cuerda, corcho,lino, galleta, algodón y madera) con 108 imágenes para cuatro muestras diferentes de cada material, en total 432 por clase. Todas las muestras fueron fotografiadas en 3 posiciones diferentes, con 4 condiciones de iluminación y a 9 escalas. Esto confiere una gran dificultad al conjunto de imágenes. En la figura 3 se pueden ver varios ejemplos de las texturas bajo diferentes condiciones.

Brodatz32 (Valkealahti and Oja, 1998) es un subconjunto de 32 imágenes (cada imagen de una clase) del conjunto original Brodatz. Está compuesto por 2048 sub imágenes de 64×64 píxeles cada una (64 por cada clase) y comprende los siguienes subconjuntos: 16 imágenes originales, 16 imágenes rotadas de la original, 16 imágenes escaladas de la original y 16 imágenes rotadas y escaladas respecto de la original. Como paso previo a utilizar este conjunto de imágenes, todas han sido procesadas para homogeneizar el histograma del nivel de gris. En la figura 4 se puede ver un ejemplo de cada una de las 32 imágenes originales.

Figura 3: Ejemplos de algunas imágenes del conjunto de imágenes KTH TIPS2-a bajo diferentes escalas e iluminaciones. De arriba a abajo: pan marrón, algodón, madera y hojas de lechuga.

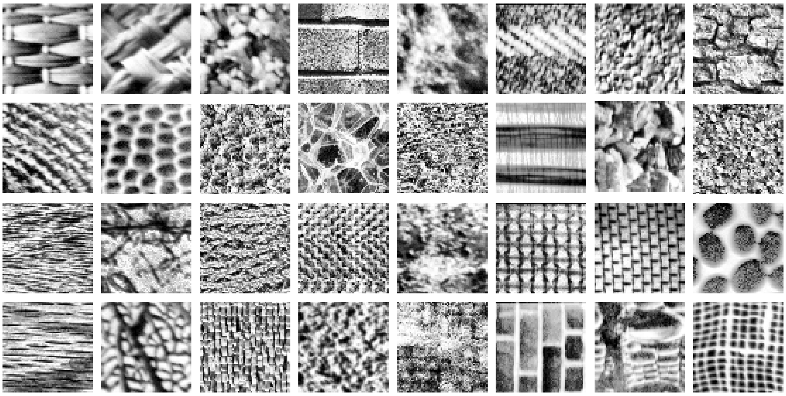


Figura 4: Ejemplos de cada clase del conjunto de imágenes Brodatz.

4.2 Conjunto de imágenes de espermatozoides

4.2.1 Vitalidad

La falta de un conjunto de imágenes público de espermatozoides vivos y muertos de verraco nos impulsó a crear este conjunto. Todas las fotografías han sido tomadas en CENTROTEC, un centro colaborador de la Universidad de Leon. 450 pares de imágenes fueron tomadas usando un microscopio de fluorescencia. Cada uno de estos pares contiene una imagen en contraste de fase positivo y una imagen de fluorescencia obtenida usando dos tinciones diferentes: propidium iodide (para colorear los muertos de rojo) y dichlorofluorescein para colorear de verde los vivos. Un ejemplo de estas capturas se puede ver en la figura 5. Una vez etiquetadas todas las imágenes se realizo un registro automático de ellas para asegurar la invarianza de escala y rotación.

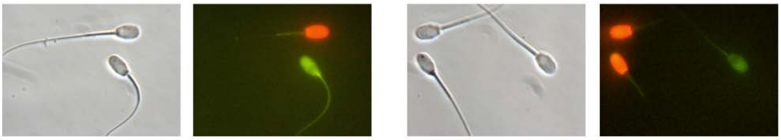


Figura 5: Dos pares de imágenes obtenidas. A la izquierda la imagen en escala de gris y a la derecha la imagen de fluorescencia.

4.2.2 Estado del acrosoma

Del mismo modo que hicimos para el conjunto de imágenes de vitalidad, hemos creado otro conjunto disponible también en nuestra pagina web[2] para determinar el estado del acrosoma. En este caso se utilizo un método automático de segmentación que permite realizar el proceso de estimación del número de espermatozoides dañados de manera totalmente automática. Finalmente se obtuvieron 1851 cabezas de espermatozoides: 905 intactas y 946 dañadas. Este conjunto de imágenes es el más extenso hasta la fecha comparado con el resto de los publicados en este tema. En solo se utilizaron 360 acrosomas mientras que en se obtuvieron 400 imágenes de cada clase. Esto implica unos resultados más reales a la hora de realizar experimentos. En las figuras 6 y 7 se pueden ver ejemplos de cabezas dañadas e intactas respectivamente.

4.3 Conjuntos de imágenes de herramientas de corte

4.3.1 Conjunto de imágenes del filo de corte

Otro campo de aplicación de la descripción de textura es la caracterización del desgaste de las herramientas de corte. Debido a la falta de conjuntos de imágenes que se ajustaran a las necesidades específicas del proyecto en el que se enmarca estos experimentos, obtuvimos el nuestro propio utilizando 53 herramientas de corte como se puede ver en la figura 8.

---

[2] http://pitia.unileon.es/VARP/datasets

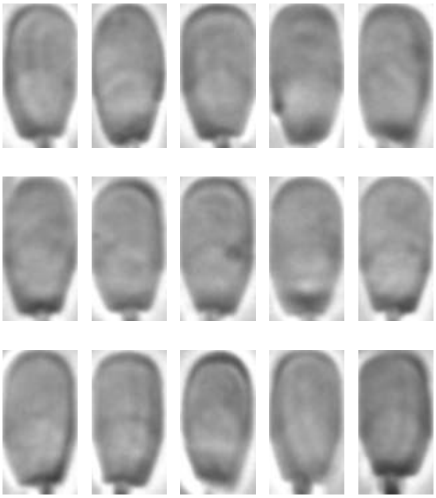


Figura 6: Cabezas de espermatozoides de verraco con el acrosoma intacto.

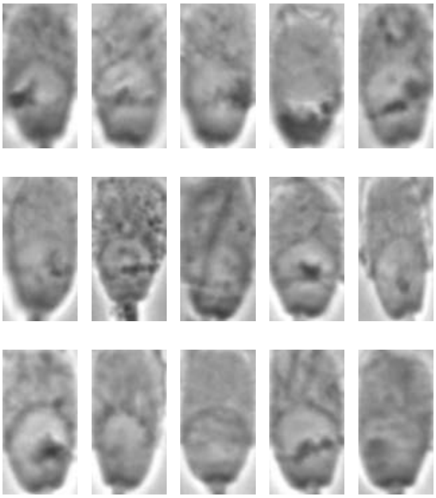


Figura 7: Cabezas de espermatozoide de verraco con el acrosoma dañado.

Las imágenes en escala de gris de las plaquitas han sido sometidas a un preprocesamiento para dividirlas en 4 sub-imágenes: una por cada filo de corte. Para ello el primer paso consiste en remover la parte central de la plaquita utilizando una mascara circular. Para determinar el centro del circulo, la imagen se binariza con un umbral de 0,01 para detectar la posición de la plaquita en la imagen respecto del fondo. Una vez detectada se calcula el centroide y se obtiene el eje de longitud ma-

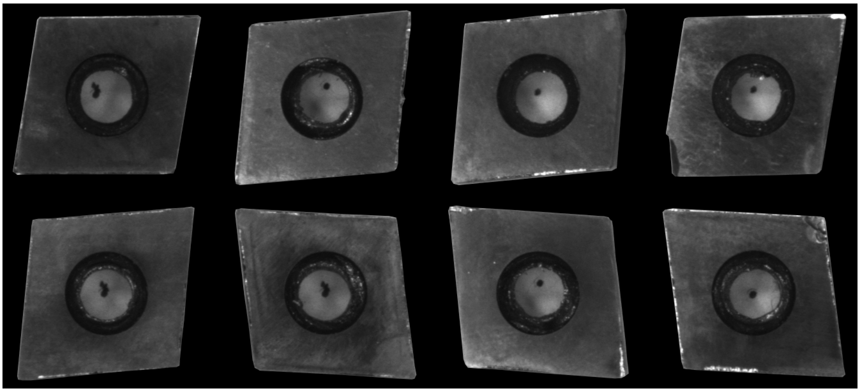


Figura 8: Ejemplos del conjunto de imágenes de plaquitas. Las plaquitas de la izquierda presentan un alto desgaste mientras que las plaquitas de la derecha tienen un bajo desgaste en relación al borde de corte izquierdo.

yor. Con esto, el radio del circulo de la región a eliminar es de 1/5 el tamaño del eje mayor. El siguiente paso consiste en aplicar un filtro Sobel a la imagen para detectar los contornos de la plaquita usando la imagen en escala de gris. Estos contornos son sometidos a un proceso de dilatación y apertura y proyectados verticalmente en el eje horizontal. El primer elemento no nulo de la proyección indica la coordenada X donde empieza la plaquita y el último la coordenada donde finaliza. Teniendo en cuenta esta posición, se realizó un recorte de 100 píxeles para recortar el filo de corte. Finalmente 212 imágenes forman el conjunto de imágenes de filos de corte. Un ejemplo de cuatro de estas imágenes puede verse en la figura 9

4.3.2 Conjunto de imágenes del desgaste del filo de corte

Para describir tan solo la región perteneciente al desgaste, se llevo a cabo una segmentación manual usando el conjunto de imágenes de filos de corte. Estas imágenes conformaron un nuevo conjunto llamado imágenes del desgaste del filo de corte. Un ejemplo de éstas se puede ver en la figura 10.

4.3.3 Conjunto de imágenes de áreas de desgaste

Las imágenes en escala de gris de las plaquitas con el fondo enmascarado fueron preprocesadas para generar una imagen por cada uno de los filos de corte. Sin embargo, en un filo se pueden identificar muchas regiones con diferentes niveles de intensidad respecto a su desgaste. Un ejemplo de la extracción del desgaste se puede ver en la figura 11. Como podemos apreciar, una imagen puede contener regiones con diferentes nivel de desgaste. Por ese motivo hemos creado un nuevo conjunto

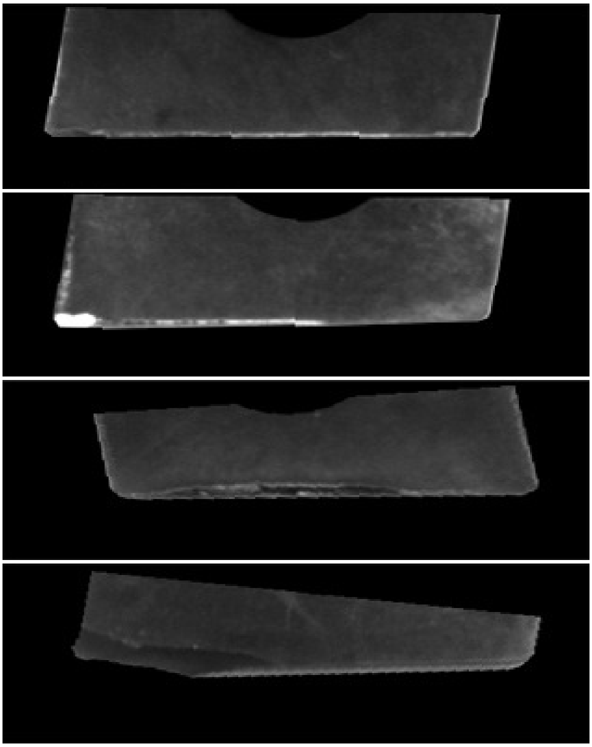


Figura 9: Ejemplo de algunas de las imágenes de filos de corte

de imágenes en el que cada una de ellas posea un único nivel de desgaste. En total, se extrajeron 577 regiones de las 212 imágenes que formaban el conjunto de datos inicial. Algunas de las imágenes de este conjunto de datos pueden verse en la figura 12. En la primera fila se muestran regiones verticales mientras que las filas dos y tres están formadas por imágenes de regiones horizontales.

4.4 Conjunto de imágenes de reconocimiento de rostros

Para el reconocimiento de rostros hemos usado el conjunto de imágenes JAFFE Lyons et al. (1998). Está compuesto de imágenes con diferentes expresiones de 10 mujeres japonesas. Todas las imágenes fueron normalizadas a 256×256 píxeles y fueron tomadas bajo las mismas condiciones de luminosidad. En nuestro caso, un preprocesamiento ha sido llevado a cabo realizando un recorte sobre e rostro para reflejar tan solo la región de la cara, evitando que el fondo influya en la descripción. El conjunto contiene 213 imágenes que representan expresiones espontáneas de miedo, disgusto, enfado, felicidad, tristeza, sorpresa y neutralidad. Un ejemplo de estas imágenes se puede ver en la figura 13.

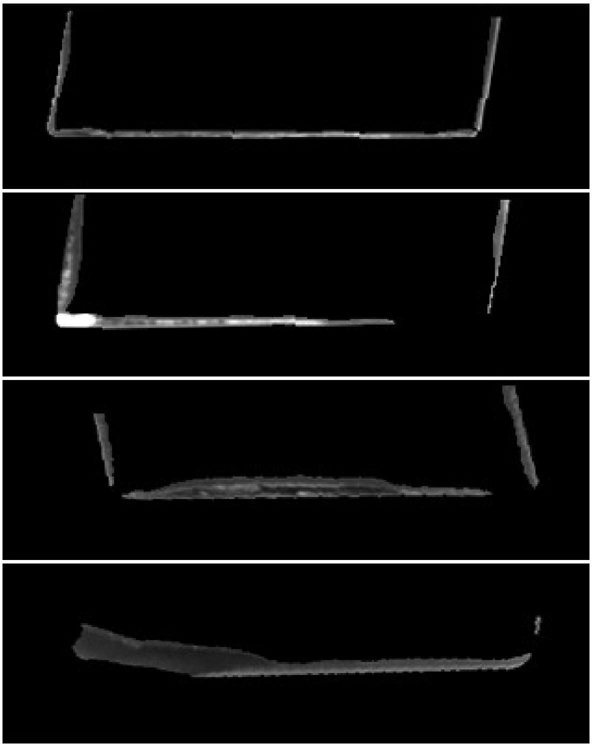


Figura 10: Ejemplos de algunas imágenes del conjunto del desgaste del filo de corte.

4.5 Conjunto de imágenes de recuperación de textiles

La recuperación de textiles en entornos reales es un campo aun por investigar. Hasta lo que nosotros sabemos, no existe ningún conjunto de imágenes público que se centre en el reconocimiento de tejidos rígidos o flexibles, presentados en diferentes tamaños, formas y condiciones de iluminación. Por esta razón, hemos creado nuestro propio conjunto para la recuperación de textiles en habitaciones, que está disponible en nuestro sitio web [3].

Está compuesto de 690 imágenes obtenidas de 15 vídeos de Youtube. Todos los vídeos han sido grabados en habitaciones repletas de diferentes textiles y desde diferentes ángulos, condiciones de iluminación, con oclusiones, etc. lo que hace que la recuperación de los tejidos sea más complicado. El conjunto de imágenes contiene 67 clases de texturas como cortinas, alfombras, sofás, camisetas, vestidos, etc. En la figura 14 se muestra un mosaico con una imagen de cada categoría. El número de elementos de cada una de las categorías varía desde 4 hasta más de 100. En total 1787 regiones han sido etiquetadas.

---

[3] http:// pitia.unileon.es

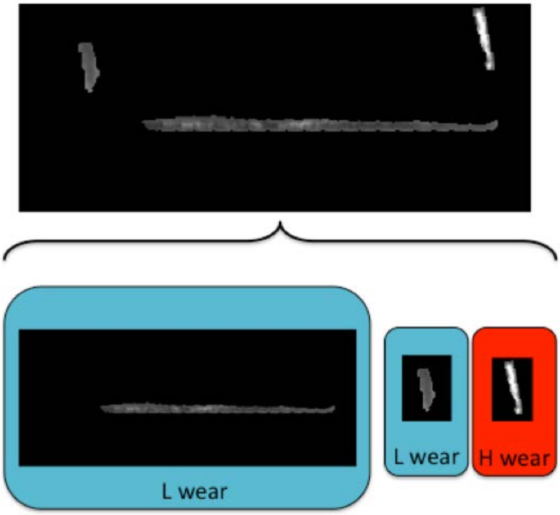


Figura 11: Ejemplo del filo de corte dividido en regiones.

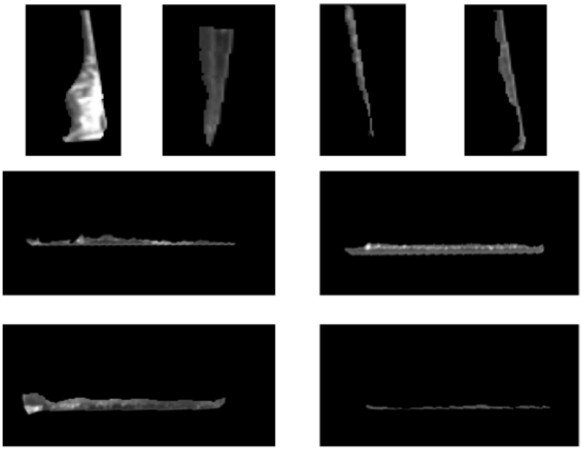


Figura 12: Ejemplo del conjunto de imágenes de regiones de desgaste.

Además de obtener las imágenes, hemos generado un XML que almacena el *Ground truth* de las imágenes. Este XML contiene las coordenadas del contenedor de la textura y la clase a la que pertenece para cada una de las imágenes. Podemos apreciar la gran diversidad que hay en términos de tamaño, posición o iluminación en los textiles en la figura 15.

Figura 13: Ejemplo de diferentes imágenes de JAFFE. En cada fila una mujer diferente y en cada columna una expresión.

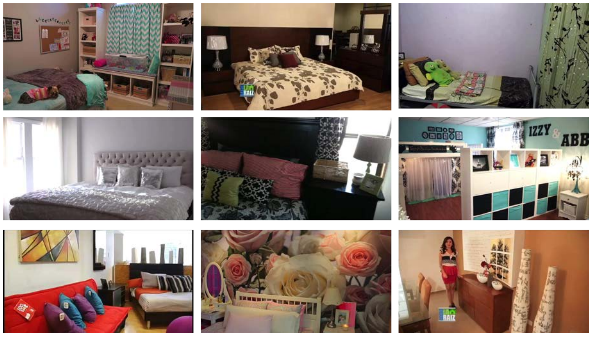


Figura 14: Una imagen de cada clase del conjunto de datos de TextilTube.

5 Adaptive Local Binary Patterns con desviación estándar orientada (ALBPS)

5.1 Method

En Guo, Zhang, Zhang and Zhang (2010) se propone el uso de la media y desviación estándar de los gradientes orientados para mejorar el proceso de clasificación. Sin embargo, su método (ALBP) no tiene en cuenta esta información en el proceso de descripción. Nuestra propuesta incluye la información relativa a la desviación estándar por orientación del gradiente de la imagen en la etapa de descripción, permitiendo una mayor flexibilidad en la clasificación y lo hemos llamado ALBPS. El vector de desviaciones estándar orientadas σ se obtiene usando la ecuación 15.

$$\sigma_p = \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{M}(g_c(i,j) - g_p(i,j) - \mu_p)^2/(M \cdot N)} \tag{15}$$

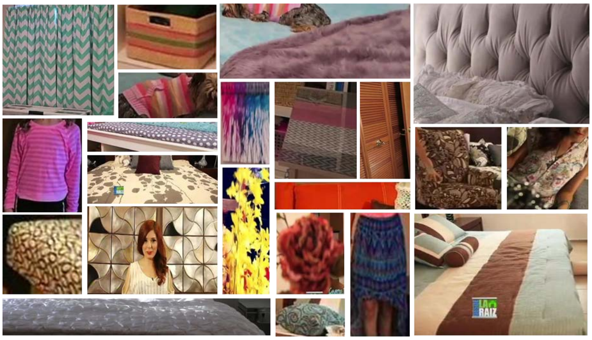


Figura 15: Ejemplos de Ground truth de regiones del conjunto de imágenes de TextilTube

donde N y M son los números de las filas y las columnas respectivamente,$g_c(i,j)$ es el píxel central de la posición (i,j), $g_c(i,j)$ es el vecindario de $g_c(i,j)$ con orientaciones $2p\pi/P$ y radio R, y μ_p es la media orientada calculada de la siguiente forma:

$$\mu_p = \sum_{i=1}^{N}\sum_{j=1}^{M} |g_c(i,j) - g_p(i,j)|/(M \cdot N) \tag{16}$$

ALBPS se obtiene al concatenar los $P+2$ valores del histograma del LBP uniforme junto con el vector de dimensión P de las desviaciones estándar, creando un vector de dimensión $2P+2$ donde P es el tamaño de nuestro vecindario.

5.2 Reconocimiento de materiales

5.2.1 Configuración del experimento

En esta sección vamos a evaluar el rendimiento de ALBPS en el conjunto de imágenes KTH TIPS2-a que está centrado en el reconocimiento de materiales, como explicamos en capítulos previos. La configuración del experimento sigue el estándar introducido por Caputo y su equipo, ya que es la más utilizada para comparar resultados con otros investigadores (Caputo et al., 2005; Chen, Shan, He, Zhao, Pietikainen, Chen and Gao, 2010). Principalmente, consiste en coger una de las muestras de cada material para probar el modelo y las otras tres para entrenarlo. Este método supone una dificultad añadida ya que en la etapa de entrenamiento nunca se utilizan imágenes de la muestra que se va a probar. Hemos usado tres aproximaciones del k vecinos más cercanos (kNN): kNN clásico, kNN con distancia media y kNN ponderado. El resultado final es la media de los ratios de acierto obtenidos usando cada muestra una vez como conjunto de prueba.

5.2.2 Evaluación de variantes de Nearest Neighbours

Los tres clasificadores basados en kNN se han utilizado para comparar diferentes versiones de LBP como el propio LBP, ALBP, LBPV, CLBP y nuestra propuesta (ALBPS). Cuatro distancias se han tenido en cuenta para realizar los cálculos: euclidea, *intersect distance*, *Chi square* y Cityblock. En la figura 16 podemos observar el rendimiento medio para los valores de k ($k = 1, 3, 5, 7, 9, 11$) de todos los clasificadores para cada par de descriptor LBP y distancia. Como podemos ver, los mejores resultados se obtienen usando el kNN ponderado para todos los casos excepto para LBP- Chi square donde el método kNN clásico es mejor.

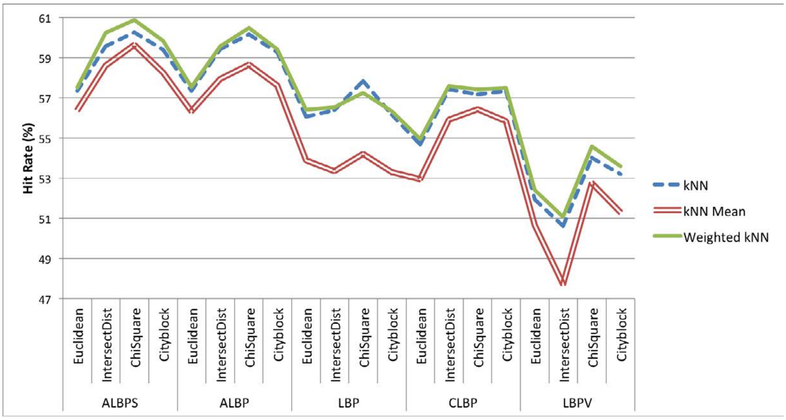


Figura 16: Resultados obtenidos en KTH tips2-a usando diferentes clasificadores. La tasa de acierto correspnde con el valor medio de las tasas de acierto obtenidas para diferentes valores de k $k = (1, 3, 5, 7, 9, 11)$ para todos los pares descriptor-distancia.

5.2.3 Comportamiento de variantes de LBP usando Weighted NN

En la figura 17, podemos observar el comportamiento medio para diferentes valores de k del kNN ponderado usando diferentes métricas de distancia. Se puede apreciar que el método propuesto, ALBPS, mejora a todos los demás para todas las distancias excepto para la euclídea, donde empata con ALBP. El mejor resultado de nuevo se obtiene usando la distancia Chi square con ALBPS con un 60.23 % de tasa de acierto

5.2.4 Mejor valor del número de vecinos (k)

Como se ha demostrado el mejor resultado lo hemos obtenido con ALBPS y la distancia Chi Square. Sin embargo, este resultado es el valor medio para diferentes valores de vecinos. En la figura 18, podemos ver el rendimiento de ALPS para los

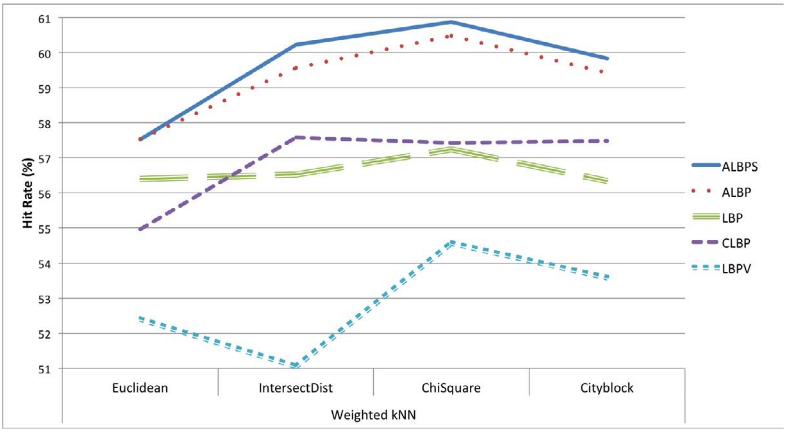


Figura 17: Resultados obtenidos en KTH tips2-a usando diferentes descriptores y el clasificador de kNN ponderado. La tasa de acierto correspnde con el valor medio de las tasas de acierto obtenidas para diferentes valores de k $k = (1, 3, 5, 7, 9, 11)$ para todos los pares descriptor-distancia.

diferentes valores de k y todas las distancias evaluadas. Los resultados muestran que el mejor porcentaje de acierto se obtiene con la distancia Chi Square para k=1 y es igual a 61.47 %.

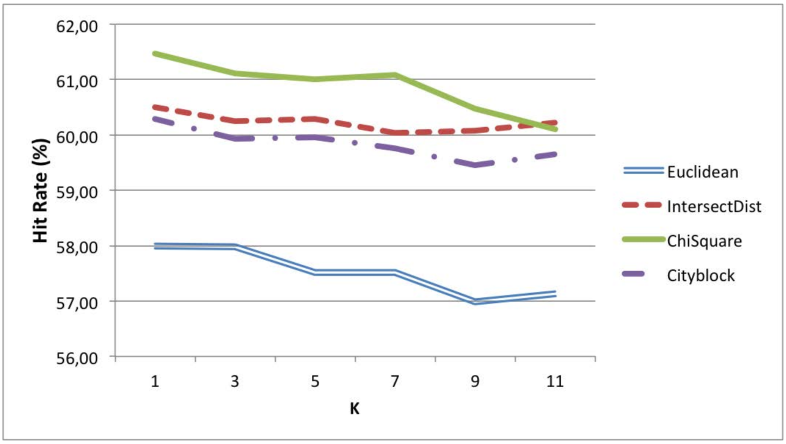


Figura 18: Resultados obtenidos en KTH tips2-a utilizando kNN ponderado y diferentes valores de k y métricas de distancia.

5.3 Clasificación de espermatozoides según su vitalidad

5.3.1 Configuración del experimento

Al igual que en el caso anterior, se tuvieron en cuenta tres aproximaciones de kNN (kNN clásico, kNN con distancia media y kNN ponderado) para clasificar espermatozoides como vivos o muertos. Hemos usado el paradigma de uno contra todos que consiste en clasificar cada imagen teniendo en cuenta la distancia a todas las demás del conjunto de imágenes. Este proceso se ha repetido para diferentes tamaños del vecindario (1, 3, 5, 7, 9 y 11).

Además de kNN también hemos utilizado un método de clasificación supervisada: máquinas de vector de soporte (SVM) utilizando *Least Squares (LS)* como método de entrenamiento y un kernel lineal. Se probaron otros kernels pero los resultados fueron peores que los obtenidos con el que se muestra. Como este conjunto de imágenes cuenta con más elementos de una clase (vivos) que de la otra (muertos) hemos usado la métrica de calidad FScore que evita resultados engañosos que puede producir el uso de la tasa de acierto. FScore se calcula como $F - Score = 2 \cdot Precision \cdot Recall/(Precision + Recall)$, donde la precisión es $Precision = TP/(TP + FP)$ y el Recall como $Recall = TP/(TP + FN)$, siendo TP el numero de verdaderos positivos, TN el número de verdaderos negativos, y FP y FN el número de falsos positivos y falsos negativos respectivamente. En este caso la clase positiva se corresponde con la de los espermatozoides muertos, al ser la menos numerosa. F-Score devuelve un resultado en el rango $[0, 1]$ donde valores cercanos a 0 indican una clasificación pobre y valores próximos a 1 una clasificación buena.

5.3.2 Evaluación de variantes de Nearest Neighbours

Comparamos nuestra propuesta ALBPS con el método original ALBP y el LBP. En la figura 19, podemos observar los resultados obtenidos usando ALBP y ALBPS con tres distancias: Euclídea, Chi square y Cityblock. En este experimento hemos excluido la distancia Intersect por su bajo rendimiento. Como podemos ver, en la clasificación de la vitalidad, el kNN ponderado no mejora al kNN clásico mientras que el kNN con distancia media obtiene siempre los peores resultados. Como el kNN original tiene un coste computacional menor, hemos elegido ese como el óptimo.

5.3.3 Comportamiento de las variantes LBP con KNN

El método propuesto ALBPS ha sido comparado con ALBP y el método original de LBP para evaluar su comportamiento. En la figura 20, podemos ver los diferentes resultados obtenidos usando kNN y diferentes métricas de distancia. ALBPS mejora

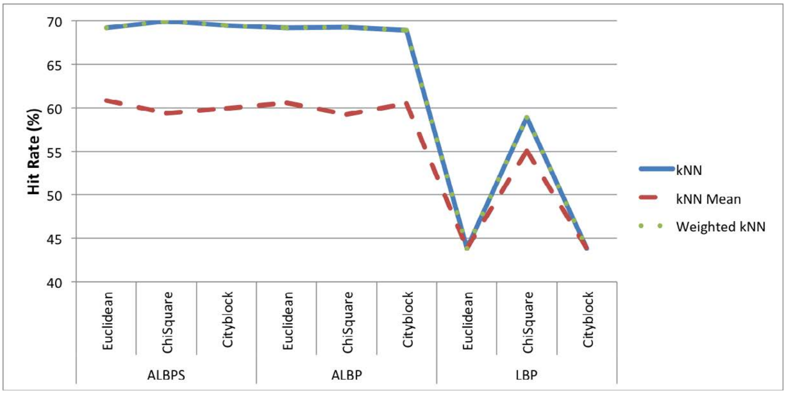


Figura 19: Resultados obtenidos en el conjunto de imagenes de espermatozoides usando diferentes descriptores y métodos kNN. La tasa de acierto corresponde con la media obtenida par alos diferentes valores de k (1, 3, 5, 7, 9, 11) usando diferentes distancias.

al resto de métodos en todos los casos, obteniendo mejoras del 1.05 % en tasa de acierto frente a ALBP y del 18.80 % con respecto al LBP. Además, esta figura muestra que, al igual que en el caso del reconocimiento de materiales, la distancia Chi square es la que obtiene el mejor comportamiento.

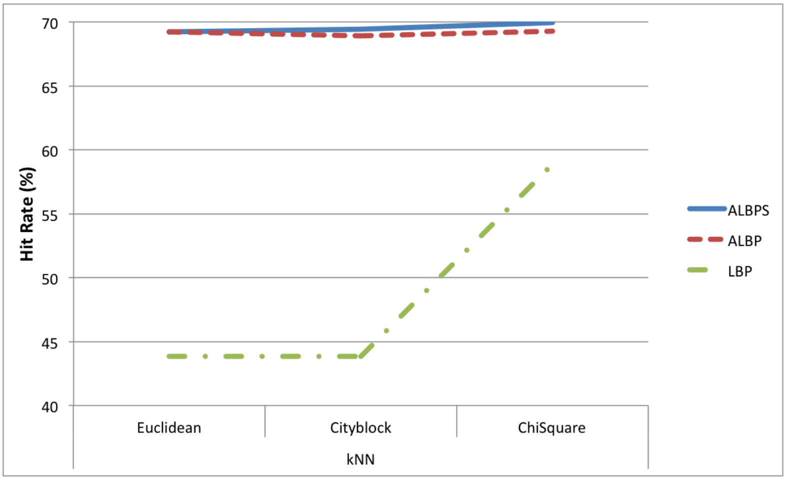


Figura 20: Resultados obtenidos con el conjunto de datos de espermatozoides usando ALBP, ALBPS u LBP.

5.3.4 Mejor valor de k

En la figura 21, mostramos el comportamiento de ALBPS para los distintos valores de k y las distintas distancias evaluadas. Los resultados muestran que el mejor valor se obtiene con la distancia Chi square para k=9 siendo este un 72.66 % de tasa de acierto.

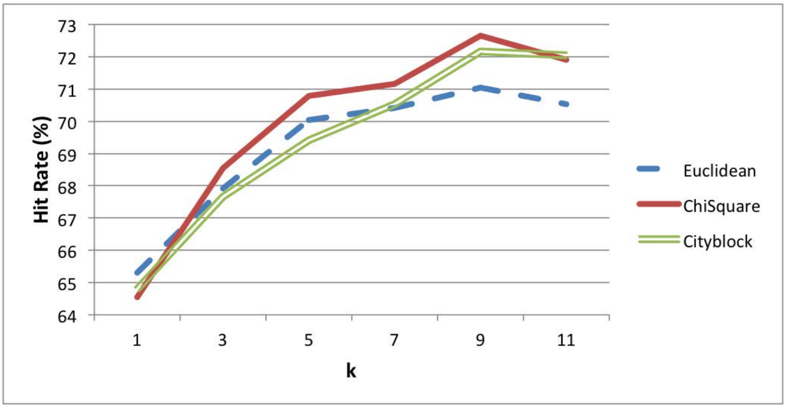


Figura 21: Resultados obtenidos en el conjunto de datos de espermatozoides para diferentes valores de k y distancias.

5.3.5 Evaluación de descriptores globales de textura con SVM

En nuestra primera aproximación, múltiples descriptores de textura conocidos en el estado del arte han sido evaluados. En la tabla 1 se pueden ver los valores de F-Score, Precision, Recall y Accuracy obtenidos con estos descriptores.

Tabla 1: Restultados de los métodos de textura globales

	WCF13	WCF4	Statistical	Legendre	Flusser	Zernike	Hu
F-Score	**0.800**	0.795	0.783	0.780	0.739	0.736	0.720
Precision(%)	**80.43**	79.54	78.72	71.00	75.58	71.38	66.24
Recall(%)	**79.57**	79.54	78.10	86.59	72.65	76.29	78.76
Accuracy(%)	**76.88**	76.75	75.69	75.00	70.88	70.19	65.44

Como se puede ver, usando WCF13 los resultados obtenidos son los mejores en comparación con el resto de métodos, obteniendo el mejor valor de FScore y Accuracy. Por contra, los resultados de métodos como el de Hu, Zernike y Flusser son muy bajos, obteniendo Hu tan solo un 65 % de precisión. Con todo esto, podemos concluir que los descriptores de textura global ofrecen resultados pobres para la clasificación de la vitalidad de los espermatozoides.

5.3.6 Evaluación de descriptores locales de textura con SVM

En este experimento, hemos usado dos descriptores de textura locales: LBP y la versión adaptativa ALBP. Su rendimiento puede verse en la tabla 2.

Tabla 2: Rendimiento de LBP y ALBP usando diferentes vecindarios.

	ALBP2,16	ALBP1,8	LBP1,8	LBP2,16
F-Score	**0.737**	0.683	0.674	0.603
Precision(%)	**73.26**	67.47	67.05	60.00
Recall(%)	**74.12**	69.14	67.82	60.67
Accuracy(%)	**71.88**	67.50	64.38	55.63

ALBP se comporta mejor que LBP en todos los casos. El mejor resultado lo obtiene ALBP2,16 con un F-Score igual a 0.737 y una tasa de acierto del 71.88 %, mejorando en más de un 15 % la precisión del LBP2,16. Aun así, ALBP no mejora los resultados obtenidos por WCF13.

5.3.7 Comparación entre ALBPS y los métodos de descripción local de textura previos

En este experimento, hemos evaluado el rendimiento de nuestra propuesta ALBPS comparando sus resultados con los obtenidos por LBP y ALBP. Además se han calculado los resultados usando el método LBP concatenado con el vector de desviaciones estándar (LBPS). En la tabla 3 y la figura 22, se muestra la comparación entre los métodos. Añadir la información de la desviación estándar provoca un incremento del rendimiento en ambos casos: LBPS y ALBPS, mejoran ambos a LBP y ALBP. El mejor resultado se obtiene con ALBPS usando$R = 2$, $P = 16$ (ALBPS2,16) obteniendo un FScore de 0.842 que supone un incremebto de un 14.25 % sobre el método original, el ALBP2,16. Además nuestro método mejora también los resultados obtenidos por el WCF13 en un 5.25 % teniendo en cuenta el FScore.

Tabla 3: Resultados obtenidos por nuestro método ALBPS y los métodos existentes.

	F-Score	Precision (%)	Recall (%)	Accuracy (%)
ALBPS2,16	**0.842**	**83.70**	**84.62**	**81.88**
ALBPS1,8	0.753	76.50	74.27	72.13
LBPS2,16	0.747	73.61	76.17	70.50
ALBP2,16	0.737	73.26	74.12	71.88
LBPS1,8	0.710	71.07	71.10	69.25

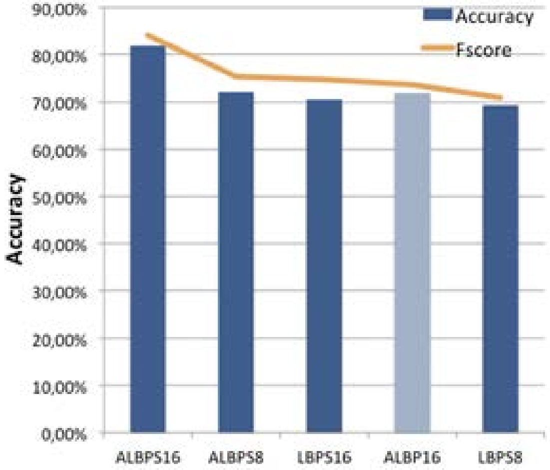


Figura 22: Resultados obtenidos con ALBPS y el mejor de los descriptores de textura locales ALBP2,16 (barra punteada).

5.3.8 Una nueva mejora: combinar características globales y locales

Nuestro último experimento consistió en combinar los métodos locales con los globales para obtener características de ambos y así describir mejor la textura de las cabezas de espermatozoides. Por lo tanto WCF13 y WCF4 fueron fusionados con los métodos locales evaluados. Los resultados se pueden ver en la tabla 4. El mejor resultado se obtuvo combinando WCF13 con ALBPS2,16 alcanzando un FScore de 0.886 y una precisión del 85.63 % mejorando los resultados obtenidos de manera individual.

Además, queremos recalcar que nuestro método híbrido WCF13+ALBPS2,16 clasificado con SVM obtiene el mejor resultado hasta la fecha, mejorando el 76.80 % de acierto obtenido en Alegre, Garcia-Olalla, Gonzalez-Castro and Joshi (2011); Alegre, Garcia-Ordas, Gonzalez-Castro and Karthikeyan (2011).

Tabla 4: Reultados de la combinación de descriptores globales y locales

	F-Score	Precision (%)	Recall (%)	Accuracy (%)
WCF13+ALBPS2,16	**0.886**	**88.12**	**89.00**	**85.63**
WCF13+LBPS1,8	0.867	89.77	84.04	85.00
WCF13+LBP1,8	0.865	93.67	80.43	85.63
WCF4+LBPS2,16	0.865	87.50	85.56	85.00
WCF13+ALBPS1,8	0.860	84.21	87.91	83.75
WCF13+LBPS2,16	0.857	80.41	91.76	83.75
WCF13+ALBP2,16	0.854	86.42	84.34	85.00
WCF4+ALBPS1,8	0.852	87.80	82.76	84.38
WCF13+ALBP1,8	0.850	85.87	84.04	82.50
WCF4+LBP1,8	0.847	86.02	83.33	81.88
WCF13+LBP2,16	0.845	82.83	86.32	81.25
WCF4+ALBPS2,16	0.843	84.27	84.27	82.50
WCF4+LBPS1,8	0.832	81.19	85.42	79.38
WCF4+ALBP2,16	0.825	82.02	82.95	80.63
WCF4+ALBP1,8	0.818	76.29	88.10	79.38
WCF4+LBP2,16	0.806	79.80	81.44	76.25

6 Local Oriented Statistical Information Booster (LOSIB)

6.1 Metodología

El principal objetivo del Local Oriented Statistical Information Booster (LOSIB) es potenciar el rendimiento de un descriptor de textura.

El concepto básico es añadir información local orientada utilizando estadísticas calculadas a lo largo de todos los píxeles de una imagen. Esta información no se suele tener en cuenta a la hora de describir texturas y aporta unos datos extremadamente útiles para la clasificación de texturas. En este trabajo, hemos llevado a cabo la combinación de LOSIB con un amplio abanico de descriptores de textura mediante la concatenación de ambos vectores. A medida que aumenta la profundidad del vecindario , LOSIB trata la información a una mayor escala. Dependiendo del problema a tratar se puede obtener mejor resultados extrayendo la información local o la global. Otro factor importante es el número de vecinos que representan las diferentes orientaciones. Si la textura es muy heterogénea, el número de vecinos debería aumentar para poder capturar toda la variabilidad de la imagen. Sin embargo, usar demasiadas orientaciones en texturas homogéneas puede ser contraproducente debido a la perdida de valor de las orientaciones realmente importante.

Por tanto, la nomenclatura de este método es LOSIB(R,P) donde R es el radio del vecindario y P el número de vecinos.

Sea c un píxel en la posición (x_c, y_c) de la imagen, p un píxel de su vecindario (con $p \in \{0, 1, ..., (P-1)\}$), cuyas coordenadas son (x_p, y_p), y sean g_c y g_p sus respectivos valores de gris, para obtener el LOSIB de una imagen lo primero es necesario extraer las diferencias absolutas d_p entre g_c y g_p,para todos los píxeles c de la imagen, siguiendo la ecuación (17). En la figura refLOSIBextractionESP se puede ver un ejemplo de la obtención de las diferencias orientadas para tres píxeles.

$$d_p(x_c, y_c) = |g_c - g_p| \tag{17}$$

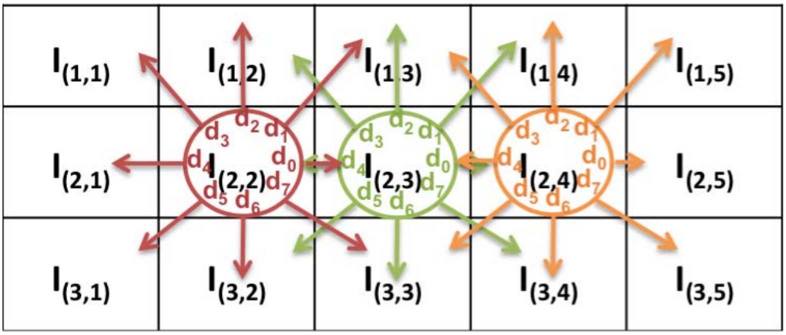


Figura 23: Ejemplo de la obtención de las diferencias orientadas para tres píxeles para calcular LOSIB(1,8).

Dado un píxel c, las coordenadas (x_p, y_p) de su p vecino se calculan usando la ecuación (18).

$$(x_p, y_p) = (x_c + R\cos(2\pi p/P),\ y_c + R\sin(2\pi p/P)) \tag{18}$$

El valor de los vecinos que no están centrados en un píxel se estima mediante interpolación de los píxeles conectados.

A continuación se calcula la media de todas las diferencias a lo largo de la misma orientación siguiendo la fórmula (19), donde N y M son el número de filas y de columnas de la imagen respectivamente.

$$\mu_p = \frac{\sum_{x_c=1}^{M} \sum_{y_c=1}^{N} d_p(x_c, y_c)}{M \cdot N} \tag{19}$$

En la figura 24, se muestra el histograma de todas las diferencias absolutas para la orientación $p = 0$ y el valor μ_0 para una imagen de KTH TIPS2-a.

Por lo tanto, LOSIB tendrá tantas características como vecinos tenga el vecindario considerado y representa el valor medio de las diferencias para todas las orientaciones mostradas en la ecuación (20).

$$\text{LOSIB}(R, P) = \bigcup_{p=0}^{P-1} \mu_p \tag{20}$$

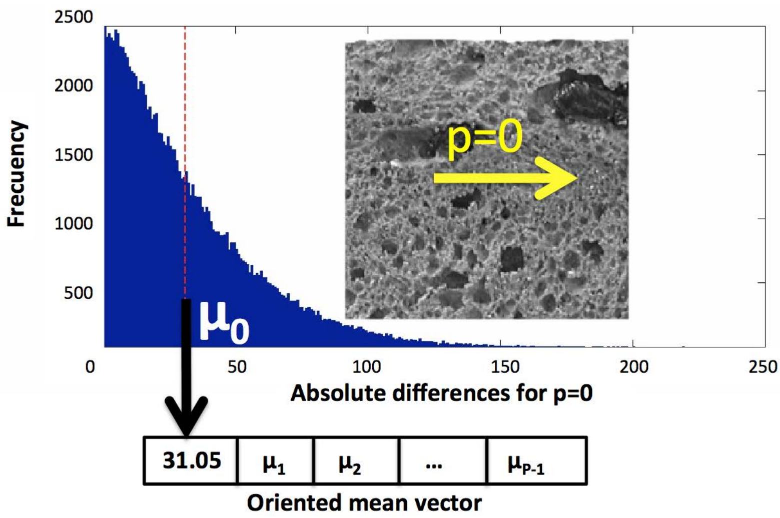


Figura 24: histograma de todas las diferencias absolutas para la orientación $p = 0$ y el valor μ_0 para una imagen de KTH TIPS2-a

6.2 Reconocimiento de materiales

6.2.1 Configuración del experimento

La configuración de los experimentos para el caso de reconocimiento de materiales es igual a la utilizada en el capítulo anterior y sigue las directrices marcadas por Caputo y su equipo de trabajo(Caputo et al., 2005; Chen, Shan, He, Zhao, Pietikainen, Chen and Gao, 2010). En el caso de Brodatz32 se ha optado por una validación cruzada de 4 repeticiones para evitar posibles efectos derivados de la aleatoriedad y se ha mostrado la media de todas ellas como el resultado final. Para ello se ha dividido el conjunto de imágenes original en un 75 % para entrenamiento y el resto para las pruebas. Se ha usado un SVM con kernel polinomial de orden dos y un paradigma de uno contra uno debido a que se trata de una clasificación multi-clase.

6.2.2 Resultados con el conjunto de imágenes KTH TIPS2-a

El primer experimento trata la sobre la clasificación del conjunto de imágenes KTH TIPS2-a usando un paradigma uno contra uno. En el se demuestra que el potenciador LOSIB mejora el rendimiento sobre todos los descriptores clásicos. Tres configuraciones diferentes de LOSIB se han evaluado: LOSIB(1,8), LOSIB(2,16), y la concatenación de ambos (que llamaremos LOSIB(1,8)+LOSIB(2,16)). En la figura 25 se muestran los resultados usando descriptores globales clásicos.

Es muy destacable el caso de los descriptores Hu, Legendre, Flusser y Zernike cuya tasa de acierto es muy pobre pero que al ser combinados con LOSIB aumenta en más de un 50 % en todos los casos (161 % en el caso de Hu.) El mejor resultado es del 63.33 % de tasa de acierto obtenido por Haralick y LOSIB(1,8)+LOSIB(2,16).

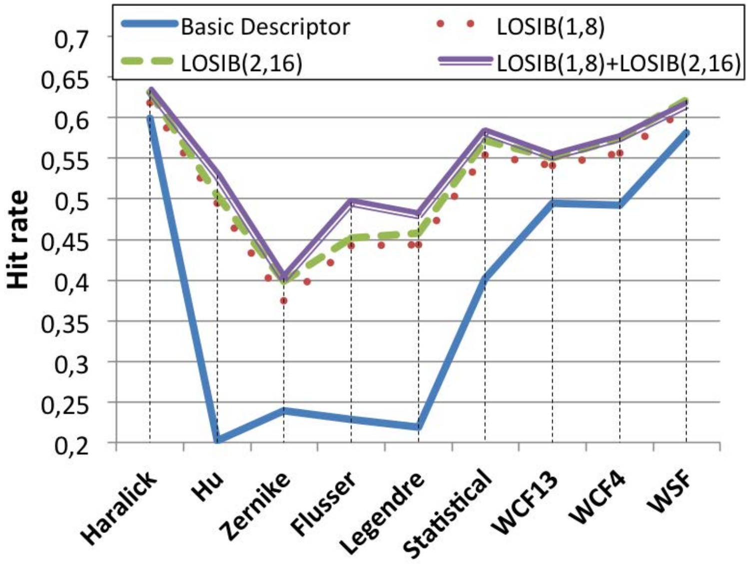


Figura 25: Resultado en KTH-TIPS 2a usando descriptores globales en combinación con LOSIB(1,8), LOSIB(2,16) y una fusión de ambos.

En la tabla 5 se muestra el porcentaje de acierto obtenido en todos los experimentos evaluados combinando LOSIB con métodos basados en LBP. El mejor acierto fue del 71.44 %, logrado por CLBP y LOSIB(1,8), pero la mejora más alta fue la conseguida al combinar LOSIB(1,8) y LBPV (8.28 % ode tasa de mejora).

El buen comportamiento de LOSIB se confirma en los experimentos llevados a cabo usando el clasificador kNN. De nuevo, el porcentaje de acierto de todos los descriptores se ve incrementado cuando son combinados con los tres potenciadores (LOSIB(1,8), LOSIB(2,16) y LOSIB(1,8)+LOSIB(2,16)). En la figura 26 podemos ver los resultados usando kNN con Chi Square como métrica de distancia. La mayor mejora en terminos de tasa de acierto se da con los descriptores globales mientras que el mejore resultado se obtuvo combinando ALBP con LOSIB(1,8)+LOSIB(2,16) con un 62.10 %.

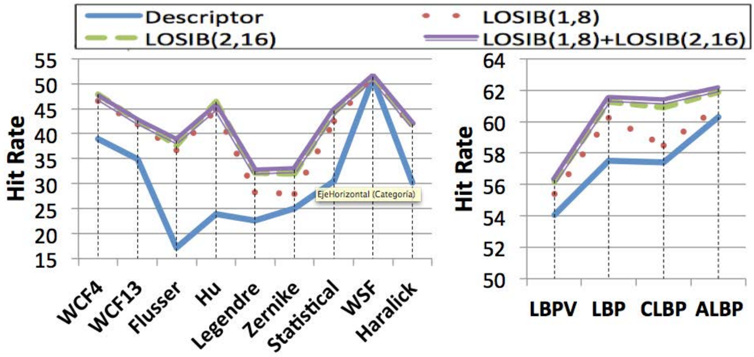


Figura 26: Resultados usando kNN en KTH-TIPS 2a. A la izquierda los descriptores globales y a la derecha los basados en LBP.

Tabla 5: Tasa de acierto, en %, de la combinación de LOSIB con descriptores globales y locales en KTH tips2-a usando SVM.

Descriptor	Haralick	Hu	Zernike	Flusser	Legendre	Statistical	WCF13
Alone	59.95±2.57	20.22±2.97	23.88±1.54	22.87±5.91	21.86±4.76	40.13±5.06	49.56±4.23
LOSIB(1,8)	61.78±5.51	49.49±5.67	37.42±4.19	44.19±5.37	44.30±7.03	55.37±6.98	54.08±3.76
LOSIB(2,16)	63.15±4.37	50.40±4.97	39.84±3.34	45.16±4.36	45.79±6.05	57.15±6.18	55.07±4.10
LOSIB(1,8)+LOSIB(2,16)	**63.32**±2.96	**52.90**±4.16	**40.13**±4.13	**49.60**±4.14	**48.02**±6.75	**58.22**±3.31	**55.28**±3.37

Descriptor	WCF4	WSF	LBP	ALBP	CLBP	LBPV
Alone	49.18 %±4.39	58.14±2.98	65.53±3.61	65.97±5.63	67.53±3.17	62.27±5.81
LOSIB(1,8)	55.68 %±6.66	61.07±4.31	65.11±7.98	66.29±7.98	**71.44**±5.16	**67.42**±8.21
LOSIB(2,16)	57.45 %±6.62	**62.16**±4.79	65.17±8.32	66.54±8.24	70.79±5.50	66.90±8.38
LOSIB(1,8)+LOSIB(2,16)	**57.51** %±5.74	61.55±4.25	**66.84**±8.00	**66.58**±8.56	70.83±5.59	67.32±6.73

6.2.3 Resultados con el conjunto de imágenes Brodatz32

Para contrastar la robustez de LOSIB en reconocimiento de materiales, se decidió realizar pruebas usando Brodatz32. Los resultados se pueden ver en la figura 27. En todos los casos, el potenciador de LOSIB mejoró los resultados de los descriptores, siendo el mejor índice de acierto un 91.06 % con LBPV y ambos LOSIB.

6.3 Clasificación de la integridad del acrosoma

6.3.1 Configuración del experimento

En la figura 28 se puede ver un esquema de la configuración del experimento. Se ha escogido un conjunto aleatorio del 20 % del total de imágenes para realizar la prueba del método. A continuación se ha llevado a cabo una validación cruzada con 10 repeticiones para escoger los mejores parámetros de configuración del clasificador SVM. El mejor modelo ha sido utilizado para la obtención del resultado final.

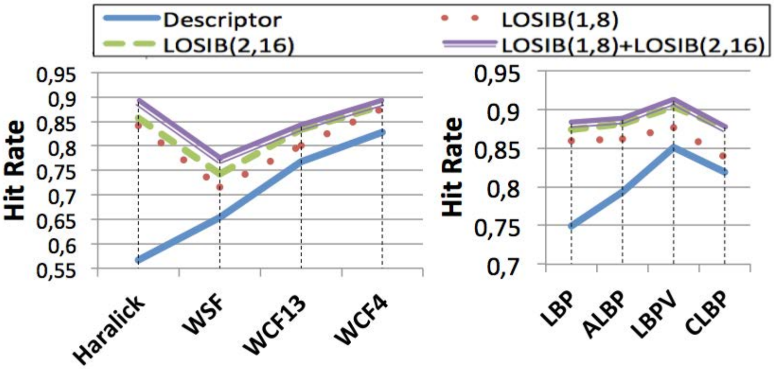


Figura 27: Resultados usando SVM en Brodatz32

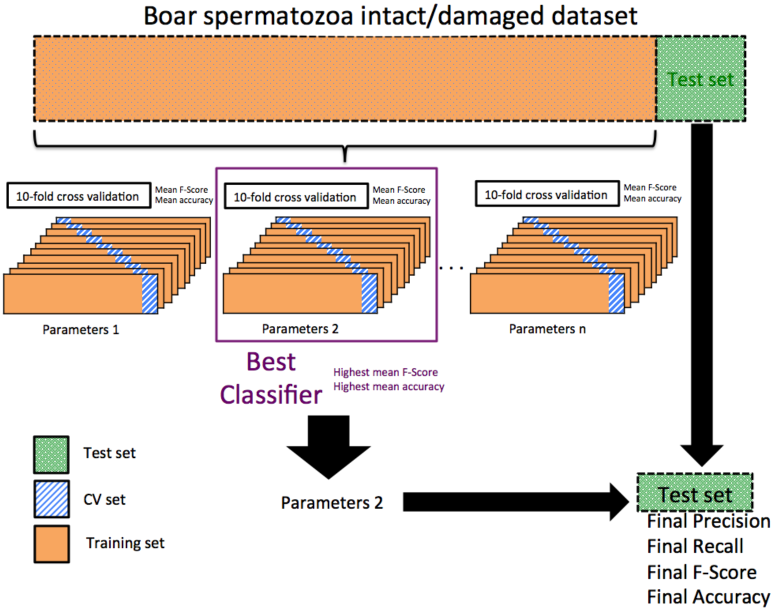
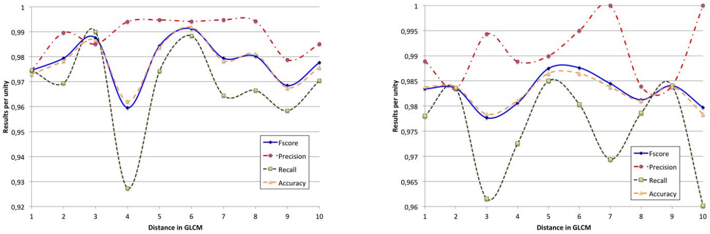


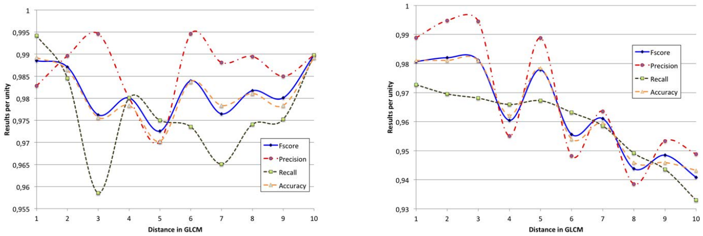
Figura 28: Esquema de la configuración de la validación cruzada.

6.3.2 Resultados

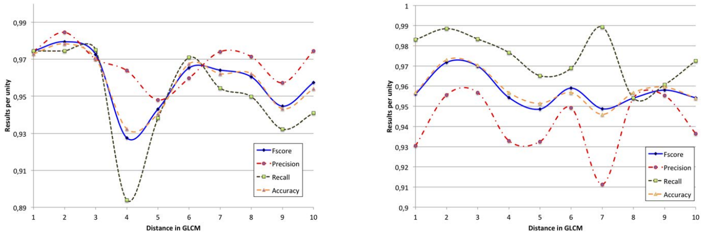
En la figura 29 se muestran los resultados para cada descriptor evaluado. WCF13 se ha calculado utilizando distancias de GLCM de 1 a 10 mientras que LBP se ha calculado con 8 (L8) y 16(L16) vecinos. Además se ha calculado un descriptor basado en el contorno de Fourier (F). Hemos considerado los siguientes descriptores: WCF13, la combinación de WCF13 con fourier (WF) o con LBP (WL8 o WL16) y finalmente la fusion de los tres métodos (WFLP8 and WFLP16). Observando la figura 29, po-



(a) WFLP8 (b) WFLP16



(c) WF (d) WL16



(e) WL8 (f) W

Figura 29: Resultados obtenidos con cada descriptor.

demos concluir que el descriptor global WCF13 (W) obtiene el mejor recall pero al mismo tiempo la peor precisión. Esto significa que casi todas las cabezas dañadas fueron etiquetadas correctamente pero que también un número importante de intactas fueron también consideradas dañadas. Al añadir la información local, *(d)* y *(e)* los resultados cambian. La información de los 8 vecinos combinados con Wavelet disminuye las diferencias entre recall y precision haciendo el método mucho más efectivo. Si observamos la subfigura *(c)*, observamos que añadiendo la información del contorno al descriptor global, la tasa de acierto aumenta. Esto demuestra que la información del contorno colabora en la identificación del estado del acrosoma. Finalmente en las subfiguras *(a)* y *(b)* se puede ver los resultados de la combinación de los tres descriptores (WCF, Fourier y LBP). Es importante el hecho de que WFLP16 obtiene un 100 % de precision. El problema de esta aproximación es que cuando la precisión es muy alta el recall cae significativamente. Por el contrario la combinación con el LBP con vecindario más pequeño (WFLP8) crea un descriptor mucho más estable con un alto FScore para distancia 3. Añadiendo LOSIB a estos descriptores el rendimiento aumenta en prácticamente todos los casos. La combinación de WFLP8 y LOSIB(1,8) disminuye el FScore pero sin embargo aumenta el recall. En cambio, la fusión de WFLP16 con LOSIB(2,16) consigue el mejor resultado obtenido hasta la fecha en problemas de clasificación de espermatozoides según el estado de su acrosoma con un FSCore de 0.9975 , una precisión del 100 % y un recall del 98.84 %. En la figura 30 podemos ver el gráfico con los datos.

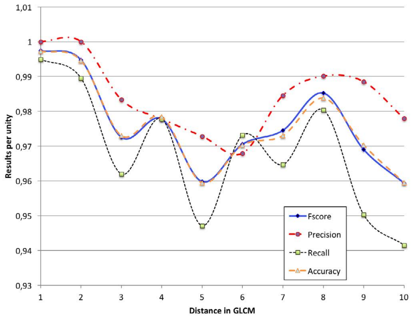


Figura 30: Resultados obtenidos por la combinación de WFLP16 y LOSIB(2,16).

Para entender mejor todos estos resultados, la tabla 6 resume todos los resultados para la mejor distancia en cada caso.

Tabla 6: Resultados obtenidos para los descriptores evaluados usando la distancia óptima en cada caso de la matriz GLCM.

	FScore	**Precision (%)**	**Recall (%)**	**Accuracy (%)**
WFLP16+LOSIB d=1	**0.9975**	**100.00**	**99.49**	**99.73**
WFLP8 d=6	0.9912	99.42	98.84	99.19
WF d=10	0.9897	98.98	98.98	98.92
WFLP16 d=6	0.9876	99.50	98.03	98.65
WL8 d=2	0.9795	98.45	97.45	97.84
WL16 d=5	0.9779	98.88	96.72	97.84
W d=2	0.9718	95.56	98.85	97.30

6.4 Estimación del desgaste en máquinas herramienta

6.4.1 Configuración del experimento

Para evitar datos incorrectos debido a la aleatoriedad, se ha empleado un algoritmo de validación cruzada utilizando la técnica de kFold con 10 repeticiones. Además se ha entrenado un SVM utilizando Least Squares y un kernel lineal. Los resultados presentados son la media de todas las iteraciones realizadas.

6.4.2 Resultados sobre el conjuntos de imágenes de filo de corte y de su desgaste

En esta sección se analizan los resultados obtenidos al clasificar las plaquitas en tres niveles de desgaste: bajo-medio-alto (L-M-H por sus siglas en inglés). El mejor resultado lo obtuvo ALBP combinado con LOSIB utilizando 8 vecinos y radio 1 con un 53.77 % de acierto. Aunque este resultado no es muy alto, este experimento nos sirve para apreciar el incremento que provoca concatenar LOSIB con los métodos de textura. En la figura 31 se pueden ver los resultados gráficamente. Además, el conjunto de imágenes del desgaste de la plaquita obtiene en todos los casos resultados mejores que el conjunto de los filos de corte.

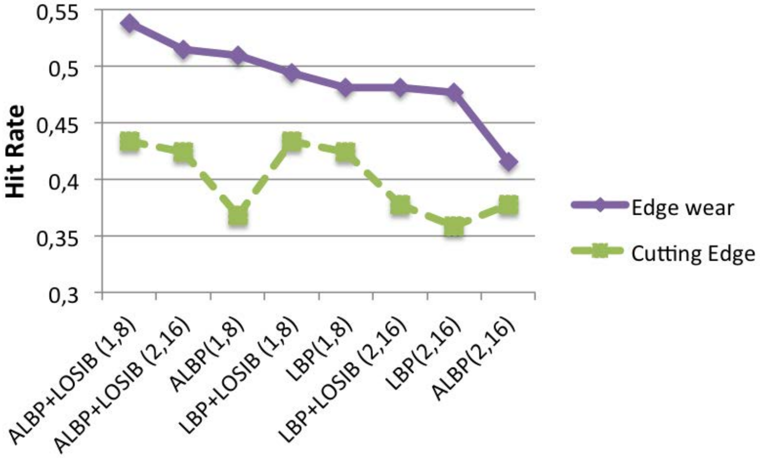


Figura 31: Resultados usando L-M-H sobre el conjunto de imágenes de filos y el de desgaste.

6.4.3 Clasificación como bajo-alto (L-H) desgaste

Los resultados obtenidos usando una clasificación binaria son mejores que los obtenidos en el experimento ternario. Para realizar este experimento, las clases baja y media se fusionaron en una sola llamada simplemente bajo, mientras que la de desgaste alto permaneció como estaba. La figura 32 muestra los resultados obtenidos. El mejor resultado lo obtuvo ALBP con LOSIB para 8 vecinos en el conjunto de imágenes de regiones al igual que en el experimento anterior. Sin embargo en este caso la tasa de acierto aumentó hasta un 74.04 % lo que supone un mejora del 38.37 % con respecto a la clasificación L-M-H.

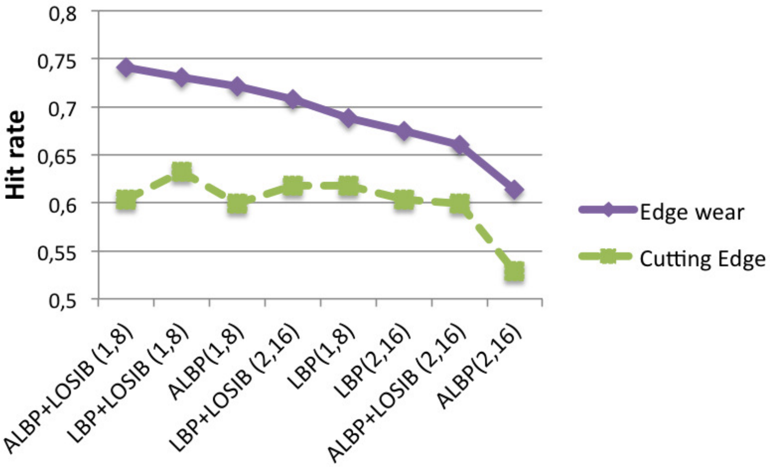


Figura 32: Resultados usando L-H sobre el conjunto de imágenes de filos y el de desgaste.

6.4.4 Resultados sobre el conjunto de imágenes de las regiones de desgaste

El primer experimento consistió en evaluar el conjunto de datos usando solo LBP y CLBP. En este caso se han omitido ALBP y LBPV debido al bajo rendimiento que obtuvieron. Además, LOSIB se ha utilizado para potenciar ambos métodos y además para describir las imágenes por si mismo.La figura 33 muestra los resultados para la clasificación binaria. Todos los métodos clásicos excepto CLBP superan a LOSIB en tasa de acierto. La figura 34, muestra los resultados para la clasificación ternaria. Se puede ver como el rendimiento usando está clasificación decrece sustancialmente. Todos los métodos obtuvieron un mejor rendimiento usando vecindarios pequeños salvo el caso del LOSIB. De nuevo LOSIB no obtiene buenos resultados de manera individual, por lo que se confirma que funciona mejor en combinación con otros métodos que por si mismo.

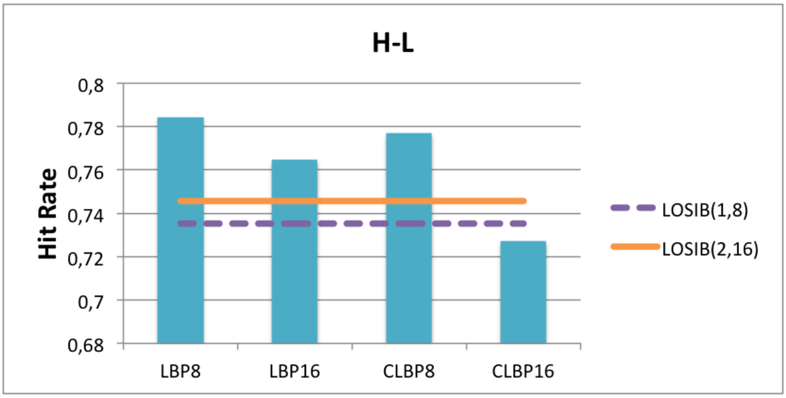


Figura 33: Resultados usando SVM con LBP y CLBP en la clasificación binaria (H-L). Las líneas horizontales muestran el valor obtenido por LOSIB.

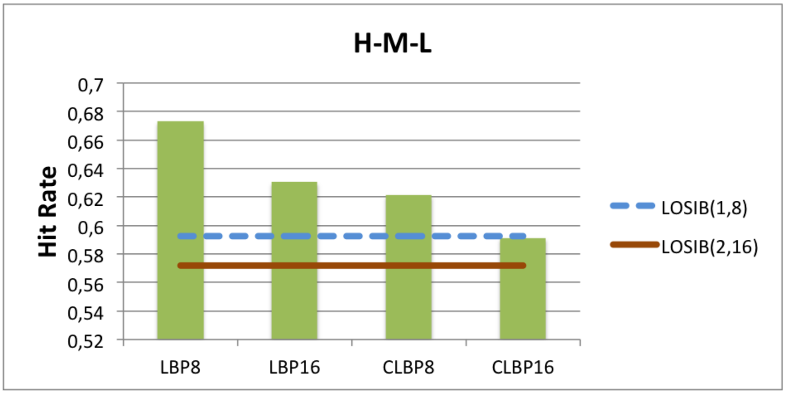


Figura 34: Resultados usando SVM con LBP y CLBP para la clasificación ternaria (H-M-L). Las líneas horizontales muestran el valor obtenido por LOSIB.

7 Complete Local Oriented Statistical Information Booster (CLOSIB)

7.1 Metodología

En este capítulo proponemos un nuevo potenciador llamado Complete Local Oriented Statistical Information Booster (CLOSIB). CLOSIB tiene como objetivo mejorar la descripción de las imágenes aportando información que otros descriptores no tratan. CLOSIB se calcula a través de la información estadística del gradiente del nivel de gris de los píxeles de una imagen. La información del gradiente de una imagen se ha utilizado en una gran cantidad de descriptores de textura en el estado de la técnica. Sin embargo, utilizar información estadística de las magnitudes del gradiente es una técnica novedosa para la descripción de imágenes. CLOSIB es conceptualmente simple y sencillo de implementar. En la ecuación 21 se define la fórmula para obtenerlo:

$$CLOSIB_{P,R,\theta} = \bigg\|_{p=1}^{P/\eta} \left((\theta-1)\mu_2^p - (-1)^\theta(\mu_1^p)^\theta\right)^{1/\theta} \tag{21}$$

donde || representa la función de concatenación, $\theta \in \{1,2\}$ es el orden del momento estadístico considerado, μ_1^p y μ_2^p son, respectivamente, el primer y segundo momento estadístico determinados por la ecuación(**??**) y η es un factor que controla la porción de orientaciones utilizadas. Si no se dice nada definimos $\eta = 1$. Por lo tanto, CLOSIB es un vector de características de dimensión P/η.

$$\mu_i^p = \frac{1}{N}\sum_{c=1}^{N}\left(d_p(x_c,y_c)\right)^i \tag{22}$$

donde N es el número de píxels en la imagen y $d_p(x_c,y_c)$ es la diferencia en valor absoluto de los niveles de gris del pixel central g_c, ubicado en (x_c,y_c), y el valor de gris del vecino g_p, ubicado a una distancia R con orientación $2\pi p/P$ respecto al píxel central.

CLOSIB permite ajustar tres parámetros: el orden del momento estadístico θ, el dario del vecindario R y la cuantización del espacio angular P.

El orden del momento estadístico, θ, determina la medida estadística usada para computar CLOSIB. Cuando $\theta = 1$, CLOSIB es un vector cuyos elementos son la media de las diferencias absolutas del pixel central con cada uno de sus vecinos. Si $\theta = 2$, CLOSIB está formado por las desviaciones estándar de las diferencias absolutas.

En relación al radio del vecindario, valores pequeños de R son muy efectivos con imágenes con un grado alto de heterogeneidad, mientras que radios grandes son buenos para imágenes con impurezas o ruido.

Respecto a la cuantización del espacio angular P, un valor alto es bueno para imágenes heterogéneas mientras que un valor bajo se ajusta mejor a imágenes con poca variabilidad en su patrón.

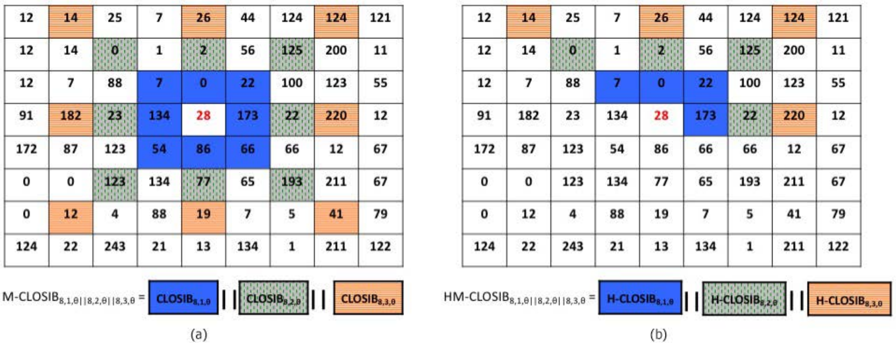
7.1.1 CLOSIB variants

En la literatura, LBP generalmente se calcula para pares de (P,R) con valores $(8,1)$, $(16,2)$ o una concatenación de ambos. Del mismo modo, CLOSIB se puede calcular mediante tripletas de (P,R,θ) igual a $(8,1,1)$, $(8,1,2)$, $(16,2,1)$, $(16,2,2)$ o como una concatenación de varias. CLOSIB se puede calcula para cualquier otro valor. El símbolo utilizado para representar la concatenación es ||. Por ejemplo, la concatenación de CLOSIB$_{8,1,1}$ y CLOSIB$_{16,2,1}$ ise representa como CLOSIB$_{8,1,1\,||\,16,2,1}$.

A continuación vamos a proponer tres variantes específicas del cálculo del CLOSIB:

CLOSIB Multiescala (M-CLOSIB) En Chan et al. (2007) se propone un método multiescala para LBP (MSLBP) utilizado en reconocimiento de rostros que aprovecha la información a multiples resoluciones para generar el histograma. MSLBP ha sido modificado y usado en un gran número de campos de la literatura (Jia et al., 2014; Wen et al., 2016).

De modo similar, hemos propuesto una variante multiescala de CLOSIB llamada M-CLOSIB. M-CLOSIB es la concatenación de CLOSIBs calculados para un determinado número de orientaciones P y múltiples radios de vecindario R. En la figura 35a se puede ver el esquema computacional para M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$.



12	14	25	7	26	44	124	124	121
12	14	0	1	2	56	125	200	11
12	7	88	7	0	22	100	123	55
91	182	23	134	28	173	22	220	12
172	87	123	54	86	66	66	12	67
0	0	123	134	77	65	193	211	67
0	12	4	88	19	7	5	41	79
124	22	243	21	13	134	1	211	122

12	14	25	7	26	44	124	124	121
12	14	0	1	2	56	125	200	11
12	7	88	7	0	22	100	123	55
91	182	23	134	28	173	22	220	12
172	87	123	54	86	66	66	12	67
0	0	123	134	77	65	193	211	67
0	12	4	88	19	7	5	41	79
124	22	243	21	13	134	1	211	122

Figura 35: Esquema que muestra los vecindarios de un pixel, marcados en rojo considerados para el cálculo de M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ (a) y HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ (b). Los píxeles marcados en azul representan el vecindario de M-CLOSIB$_{8,1,\theta}$ (a) and HM-CLOSIB$_{8,1,\theta}$ (b). Los píxeles pintados de gris muestran el vecindario de M-CLOSIB$_{8,2,\theta}$ (a) and HM-CLOSIB$_{8,2,\theta}$ (b). Los píxeles rosas determinan el vecindario de M-CLOSIB$_{8,3,\theta}$ (a) and HM-CLOSIB$_{8,3,\theta}$ (b). M-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ se obtiene como la concatenación de CLOSIB$_{8,1,\theta}$, CLOSIB$_{8,2,\theta}$ y CLOSIB$_{8,3,\theta}$. HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$ como la concatenación de H-CLOSIB$_{8,1,\theta}$, H-CLOSIB$_{8,2,\theta}$ and H-CLOSIB$_{8,3,\theta}$.

Half CLOSIB (H-CLOSIB)

Para valores pares de P, CLOSIB tiene en cuenta las diferencias absolutas de los píxeles $d_p(x_c, y_c)$ a lo largo de direcciónes que distan en π radianes.La figura 36a muestra este caso para $P = 8$. La información estadística a lo largo de direcciones que distan π radianes es generalmente muy similar. En la figura36(b-c) se pueden ver dos ejemplos de esto.

Definimos Half CLOSIB (H-CLOSIB) mediante la ecuación 21 con $\eta = 2$. El espacio angular es cuantificado en P partes pero solo se tienen en cuenta las primeras

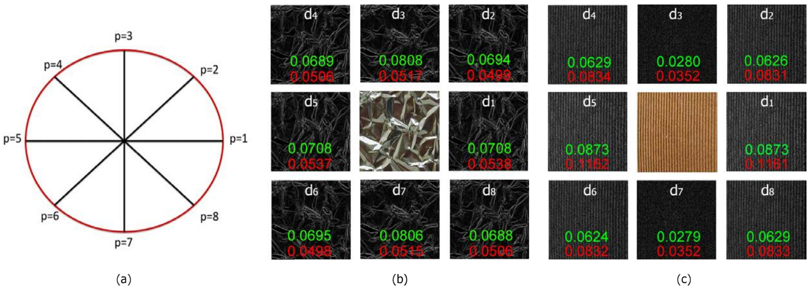


Figura 36: (a) Circunferencia que representa el vecindario considerado para el cálculo de CLOSIB con $P = 8$. Cuatro pares de vecinos difieren en π radianes. (b) y (c) representan el cálculo de CLOSIB$_{8,1,\theta}$ para dos imágenes diferentes. Podemos ver la imagen original en el centro y las ocho imágenes de las diferencias absolutas del nivel de gris en la capa exterior. Los números rojos y verdes representan, respectívamente el valor de CLOSIB$_{8,1,1}$ y CLOSIB$_{8,1,2}$ obtenidos para el correspondiente valor de p.Se puede apreciar que lso elementos ubicados en la misma dirección divergen en un máximo de 0.0002 unidades mientras que los ubicados en diferentes direcciónes en al menos 0.0006 unidadess.

P/η orientaciones. La figura 37 muestra un ejemplo de las orientaciones consideradas para calcular CLOSIB$_{8,1,\theta}$ y H-CLOSIB$_{8,1,\theta}$.

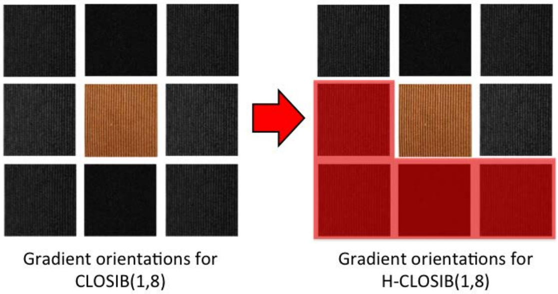


Figura 37: Esquema del cálculo de CLOSIB$_{8,1,\theta}$ (a) y H-CLOSIB$_{8,1,\theta}$ usando el ejemplo de la figura 36(c).

H-CLOSIB posee dos cualidades. La primera es que elimina información redundante del descriptor y la segunda que disminuye la dimensionalidad del CLOSIB equivalente en la mitad. Esto puede ser a veces un factor importante cuando el sistema en el que se monta el proceso de descripción tiene una memoria limitada o el tiempo es crítico.

Half Multi-scale CLOSIB (HM-CLOSIB) Half CLOSIB multiescala se obtiene cuando en M-CLOSIB el valor de $\eta = 2$. Esta variante combina las ventajas y desventajas de M-CLOSIB y H-CLOSIB. La figura 35b muestra el esquema del cálculo de HM-CLOSIB$_{8,1,\theta\,||\,8,2,\theta\,||\,8,3,\theta}$.

7.2 Reconocimiento de materiales en KTH TIPS2-a

7.2.1 CLOSIB frente a descriptores basados en LBP

Aunque CLOSIB fue creado como un potenciador de otros descriptores, en esta sección vamos a evaluar el rendimiento de CLOSIB como si fuera un descriptor por si mismo. La figura 38 muestra los resultados obtenidos cuando describimos las imágenes con CLOSIB y con descriptores basados en LBP. En todos los casos, el mejore resultado de CLOSIB se obtuvo al concatenar los valores del primer y el segundo momento estadístico.Para todos los métodos basados en LBP el mayor acierto se obtuvo con $R = 2$ y $P = 16$ vecinos. La mayor tasa de acierto se obtuvo con HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ siendo su valor 67.93 %. Esto demuestra que el método propuesto mejora pos si mismo a varios de los métodos basados en LBP de la literatura.

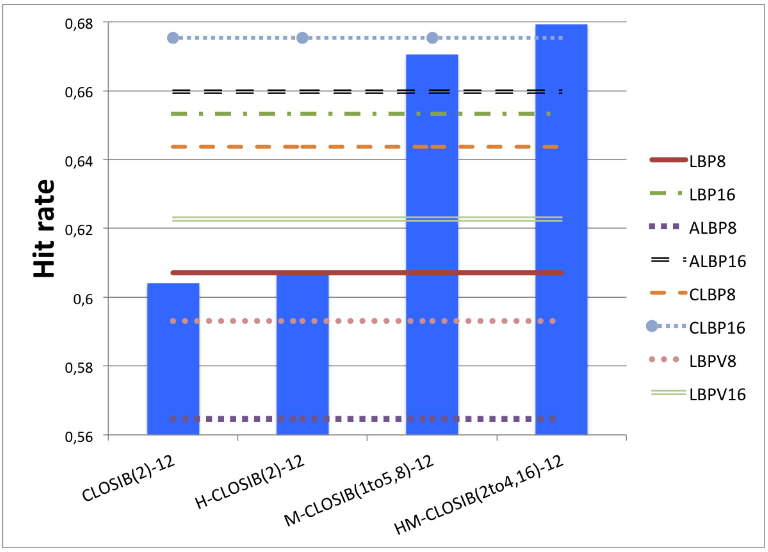


Figura 38: Tasa de acierto al describir KTH Tips2-a usando diferentes CLOSIBs y métodos basados en LBP. Para cada variante de CLOSIB, tan solo se ha representado el mejor valor obtenido para las diferentes combinaciones de parámetros.

7.2.2 CLOSIB + descriptores basados en LBP

El siguiente experimento consistió en combinar métodos basados en LBP con CLOSIB. La figura39 muestra los resultados.

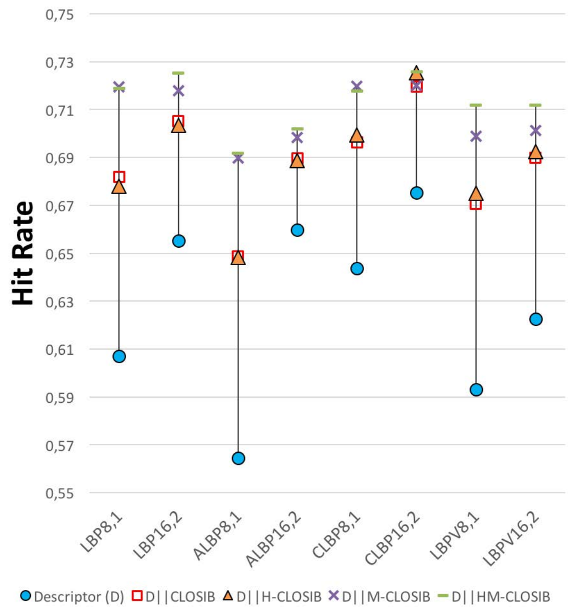


Figura 39: Tasa de acierto obtenida al combinar métodos basados en LBP con variantes de CLOSIB.

En todos los experimentos, los mejores resultados se obtienen al combinar CLOSIB con los descriptores LBP. La tasa de acierto mayor ha sido de 72.54 % usando CLBP$_{16,2}$ ||HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ and CLBP$_{16,2}$ ||H-CLOSIB$_{16,2,1\,||\,16,2,2}$ seguido de cerca de LBP$_{16,2}$ ||MH-CLOSIB$_{8,1,1\,||\,8,1,2\,||\,8,2,1\,||\,8,2,2\,||\,8,3,1\,||\,8,3,2\,||\,8,4,1\,||\,8,4,2\,||\,8,5,1\,||\,8,5,2}$ con un acierto del 72.50 %.

7.2.3 Comparativa con el estado del arte

Multitud de grupos de investigación han usado el conjunto KTH TIPS2-a para realizar sus evaluaciones. La tabla 7, muestra los resultados obtenidos en multitud de artículos desde 2010 en adelante. En todos los casos el método propuesto CLBP$_{16,2}$ ||HM-CLOSIB$_{16,2,1\,||\,16,2,2\,||\,16,3,1\,||\,16,3,2\,||\,16,4,1\,||\,16,4,2}$ mejora los trabajos previos. Además usando tan solo HM-CLOSIB como descriptor, se mejoran también un alto número de métodos.

Tabla 7: Tasa de acierto de artículos del estado de la técnica en KTH TIPS2-a.

Descriptor	**Acierto (%)**	**Año**	**Cita**
WLD	56.4	2010	(Chen, Tian, Lee, Zheng, Smith and Laine, 2010)
MWLD	64.7	2010	(Chen, Tian, Lee, Zheng, Smith and Laine, 2010)
LHS	73.0	2012	(Sharma et al., 2012)
CMLBP	73.1	2012	(Li and Fritz, 2012)
LQP	64.2	2012	(Hussain and Triggs, 2012)
CMR	69.4	2013	(Zhang et al., 2013)
PC	71.5	2013	(Zhang et al., 2013)
WLBP	64.4	2013	(Liu et al., 2013)
DRLTP	62.6	2014	(Satpathy et al., 2014)
DRLBP	59.0	2014	(Satpathy et al., 2014)
AMBP	70.3	2015	(Hafiane et al., 2015)
HM-CLOSIB	67.9		
CLBP$_{16,2}$ || HM-CLOSIB	**72.54**		

7.3 Reconocimiento facial

Dos experimentos utilizando JAFFE se han llevado a cabo para evaluar CLOSIB: Uno con las imágenes originales y otro con las imágenes recortadas.

La figura 40 muestra las tasas de acierto para el primer experimento. En todos los casos combinar CLOSIB con LBP o sus variantes mejoran al método original.

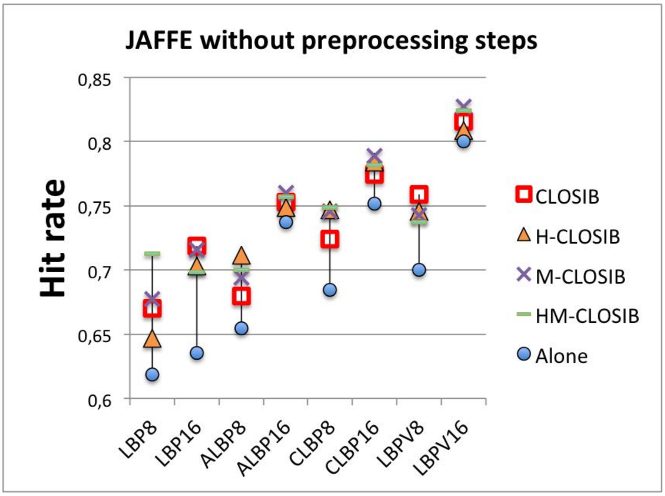


Figura 40: Resultados usando la concatenación de CLOSIB y métodos basados en LBP sobre JAFFE

Respecto al segundo experimento, la figura 41 muestra las tasas de acierto con las imágenes recortadas. De nuevo, la combinación de LBP y CLOSIB mejora los resultados, obteniendo el mayor de ellos (90.00 %) usando CLBP$_{16,2}$ || H-CLOSIB$_{16,2,1||16,2,2}$.

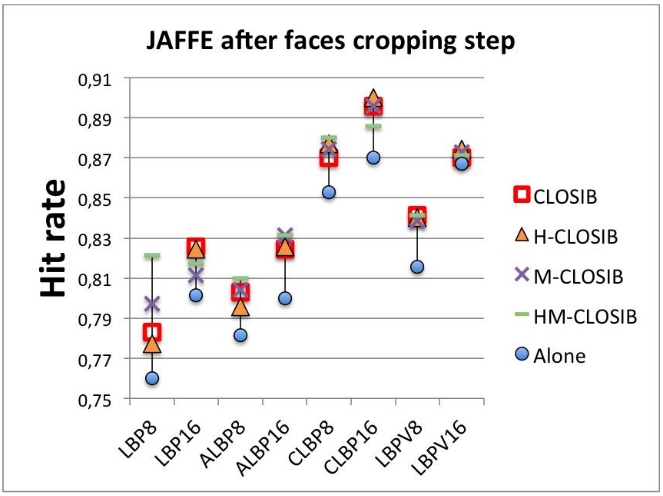


Figura 41: Resultados usando la concatenación de CLOSIB y métodos basados en LBP sobre JAFFE recortado

8 Conclusiones

En cumplimiento del punto 3º del artículo 19 del Reglamento de las enseñanzas oficiales de doctorado y del título de doctor en la Universidad de León, aprobado en Consejo de Gobierno el 25/9/2012, las conclusiones, contribuciones y trabajo futuro de esta tesis, han sido presentadas en el capítulo 9.