Chapter 1

Introduction

1.1. Motivation

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

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

- **Classification**: Given an image patch, decide which of the multiple possible categories is present in that patch.

- **Detection and localisation**: Given a complex image, decide if an specific object of interest is located somewhere in this image, and provide accurate location information on the object.

- **Content-based image retrieval**: provide automated indexing of images for their retrieval from a dataset, according to the detection and localisation of an object of interest.

This dissertation studies some particularities of object recognition through three different applications: classification of boar spermatozoa according to the acrosome...
1. Introduction

integrity; automatic identification of broken inserts in edge profile milling heads and finally retrieval of objects for the *Advisory System Against Sexual Exploitation of Children* project in relation with the evaluation of a number of clustering techniques applied to keypoint descriptors and the improvement of an existing method, COSFIRE filters, by adding the capability of describing objects using also colour information. The three aforementioned applications are highly relevant tasks that pose challenging current computer vision problems on the field of object recognition. Two key reasons led to the selection these applications. On the one hand and more importantly, these applications allow to select and to develop appropriate object description and retrieval techniques for different purposes, fields and data sets, which is the main goal of this thesis. Therefore it allows the study of object recognition from different perspectives with a knowledge transfer from one application to another. On the other hand, our research group at University of León received funds to find solutions to these applications in collaboration with companies that work on the topic. In the following, the motivation of each application is presented.

1.1.1. Classification of boar spermatozoa according to the acrosome integrity

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

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

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

This task has been studied using digital images taken on samples without staining and using a phase-contrast microscope. The existing approaches make use of
1.1. Motivation

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

standard texture description of the spermatozoa heads. These solutions need to segment the heads of the spermatozoa, extract the patterns that characterise them and classify those patterns to finally estimate the rate of damaged acrosomes present in the sample (González-Castro et al., 2009). The segmentation itself is a critical task that represents a yet unsolved problem. By using invariant local features (ILF), this segmentation step can be avoided. In this work we present several approaches where the classification of boar spermatozoa is carried out using different techniques based on ILF.

1.1.2. Localisation of broken inserts in edge profile milling heads

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

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

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

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

TECOi is a company interested in the development and installation of TWM systems that are able to automatically detect broken inserts. TECOi provided us such an edge profile milling head tool and the cutting tools to study the automatic inspection of inserts.

1.1.3. Object recognition for content-based image retrieval

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

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

Invariant local features (ILF) can rely on features clustering in order to improve the matching process. First, the matches between keypoint descriptors of the ROI image and the query image are computed. Then, we should adopt a criterion to assure if there is a real correspondence between images and, if any, the strength of that correspondence. One possibility is to use the distance of the closest match between the ILF descriptors of the ROI and the query image. Thus, the hit list would be created by sorting those computed distances, and a threshold could be set up to decide the minimum value of distance at which a correspondence is considered. However, this could lead to two kinds of errors. On the one hand, the local surroundings of two keypoints could be very similar even when they belong to different objects. On the other hand, unfortunately an ill-selected bounding box makes that the query object comes jointly, partially or completely, with other objects or cluttered background in the ROI. Lowe (2004) suggested to consider clusters of at least 3 features that agree on an object and its pose for reliable object recognition. He proposed to
use Hough transform to identify clusters that vote for the same pose of an object and to perform a geometric verification through least squares solution for consistent pose parameters. Nonetheless, there is a lack of reasoning for the choice of this clustering approach and its theoretical insight. We evaluate both approaches, direct matching to the closest pair of correspondences and the use of Hough transform with least squares verification in the scope of ASASEC project. For the latter, we compare different configurations of clustering sets of keypoints in relation with their pose parameters: coordinates location, scale and orientation obtained with scale invariant feature transform (SIFT) method.

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

1.2. Objectives

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

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

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

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

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

1.3. Main contributions

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

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

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

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

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

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

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

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

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

1.4. Thesis Organization

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

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

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

fusion of ILF with global texture descriptors for the description of the spermatozoa heads.

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

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

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