AN APPROACH TO AUTOMATIC SELECTION OF THE OPTIMAL LOCAL FEATURE DETECTOR

Bruno Ferrarini

A thesis submitted for the degree of
Master of Science by Dissertation

at the

School of Computer Science and Electronic Engineering

University of Essex

September 2016

To my grandmother Marina

05.08.1928 - 31.10.2014

Acknowledgements

I am grateful to Prof. Klaus McDonald-Maier and Dr. Shoaib Ehsan. They guided me through this course of study and giving me the opportunity to learn a lot about the art of research.

I wish to thank my group mates for their support in my academic life at the university: Hasan Tahir, Ali Khattab, Georgios Stamatiadis and Gerry Gialias.

The University of Essex meant also sport and social life to me. I want to thank my dear friends at the Archery Club for the amazing moments spent together at the competition in the University League: Hamid Jalalian, Huseyin Serkan Sahan, Kacper Radomski, Karl Wilson and our president, Saira Hussain. Also, I will remember forever the interesting discussions with George Polychronopoulos and Evangelos Stravelas on economics and finance.

Abstract

Feature matching techniques have significantly contributed in making vision applications more reliable by solving the image correspondence problem. The feature matching process requires an effective feature detection stage capable of providing high quality interest points. The effort of the research community in this field has produced a wide number of different approaches to the problem of feature detection. However, imaging conditions influence the performance of a feature detector, making it suitable only for a limited range of applications. This thesis aims to improve the reliability and effectiveness of feature detection by proposing an approach for the automatic selection of the optimal feature detector in relation to the input image characteristics. Having knowledge of how the imaging conditions will influence a feature detector's performance is fundamental to this research. Thus, the behaviour of feature detectors under varying image changes and in relation to the scene content is investigated. The results obtained through analysis allowed to make the first but important step towards a fully adaptive selection method of the optimal feature detector for any given operating condition.

Contents

List of Figures

List of Tables

Abbreviations

BRIEF	Binary Robust Independent Elementary Features
EBR	Edge-Based Region
FAST	Features from accelerated segment test
GLOH	Gradient Location and Orientation Histogram
HARAFF	Harris-Affine
HARLAP	Harris-Laplace
HESAFF	Hessian-Affine
HESLAP	Hessian-Laplace
HoG	Histogram of Oriented Gradients
IBR	Intensity-Based Region
MSER	Maximally Stable Extremal Regions
ORB	Oriented FAST and Rotated BRIEF
SALIENT	Salient Regions
SFOP	Scale-invariant Feature Operator
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Feature
SVM	Support Vector Machine

List of Publications

Journal Publications

1. **Bruno Ferrarini**, Shoaib Ehsan, Naveed Ur Rehman, and Klaus D. McDonald-Maier. "Performance comparison of image feature detectors utilizing a large number of scenes." Journal of Electronic Imaging 25, no. 1 (2016): 010501-010501.

Conference Publications

1. **Bruno Ferrarini**, Shoaib Ehsan, Naveed Ur Rehman, Ales Leonardis and Klaus D. McDonald-Maier "Automatic Selection of the Optimal Local Feature Detector." Image Analysis and Recognition (ICIAR), 2016 Internation Conference on, pp. 284-289, 2016.

2. **Bruno Ferrarini**, Shoaib Ehsan, Naveed Ur Rehman, and Klaus D. McDonald-Maier. "Performance characterization of image feature detectors in relation to the scene content utilizing a large image database." In Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on, pp. 117-120. IEEE, 2015.

3. Shoaib Ehsan, Adrian F. Clark, **Bruno Ferrarini**, Naveed Ur Rehman, and Klaus D. McDonald-Maier. "Assessing the performance bounds of local feature detectors: Taking inspiration from electronics design practices." In Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on, pp. 166-169. IEEE, 2015.

1

Introduction

This chapter discusses the motivation for this thesis and presents the most important challenges towards an effective method for selecting the optimal local feature detector in relation to the characteristics of an input image. The primary contributions of this thesis are also discussed. Finally, the organisation of the following parts of this dissertation is described.

Looking at an image and describing its content is an almost effortless task for humans. However, when it comes understanding how this process works and how to model it as a computer application, it results as an incredibly hard challenge. Computer vision discipline studies the process of acquiring, processing and analysing images in order to reproduce the human ability to see and understand the surrounding environment. Although a complete vision system capable of matching human vision it seems to be far to come, decades of research in this field made possible the creation of a wide variety of real-world applications. Just a few examples are material and weld joint inspection by means of stereo vision with X-rays images; medical diagnostics by images; stitching, which consists in turning overlapping photos into a uniform panorama; 3D object model reconstruction from one or more snapshots of a scene.

The developing of feature matching techniques greatly contributed making vision applications more robust by solving the problem of the image correspondence under image image transformations such as viewpoint and blur changes. The feature matching process can be represented with a pipeline of three independent steps. The first step detects features in an image, which are distinctive elements such as edges, corners and blobs. The second step computes descriptors for the detected features, which are utilised by the last step to perform the feature matching.

Feature matching is frequently employed in the first stage of vision applications, so it is not surprising that the overall performance of a vision system greatly depends on the quality of the feature extracted. For this reason, the research community has been very active in the last decades providing a number of different approaches to the problem of feature detection [1]. However, in spite of the significant advances achieved by far, a feature detection method that performs equally well under any imaging conditions has not been obtained yet. Indeed, it is well known that specific feature detectors perform well only in a limited range of applications [1] [2] [3]. At cost of a high computational demand, an obvious solution to this problem is running multiple feature detectors so that the shortcomings of one detector are countered by the strengths of the other detectors. This thesis presents an alternative approach based on the automatic selection

of the optimal local feature detector in relation to imaging conditions. This will allow designing adaptive vision systems capable of employing the most suitable feature detector to cope the context where they are operating.
The following sections of this chapter present the challenges towards a working and efficient method to solve the problem of the selection of the optimal feature detector followed by a summary of the main contributions provided by this research. The chapter ends with a description of the thesis structure.

1.1 Challenges

The selection of the most suitable feature detector for a particular image poses several challenges. As mentioned before, predicting how a detector will perform in a particular context is fundamental for this research thus, obtaining a reliable method to characterise a detector's performance is a primary challenge to deal with. A feature detector might be employed in several types of vision applications, operating in complex and unknown environments. However, most of the performance metrics currently available do not always reflect the actual performance of a detector [1], so utilising a suitable performance metric is the first condition to meet in order to obtain a reliable evaluation method. The proposed approach to the automatic selection of the optimal local feature detector requires having a comprehensive performance model of detectors' behaviour with a wide variety of scenes and in different imaging conditions such as brightness changes or scaling variation. Indeed, both the scene characteristics and the image transformations have an impact on a detector's performance [1] [2] [3] but a comprehensive description of how those factors combined together affect feature detectors' performance has received none or little attention so far. In such an analysis, not only the performance metric utilised but also the image datasets employed are important in order to obtain a reliable performance characterisation. In particular, a suitable image database should include a wide variety of real-world scenes and image transformations at varying rates [4] [5].
The problem of analysing a scene to determine its characteristics is another important issue to consider in the context of this research. From an image, both the image

transformation and the scene content need to be determined and understood in order to determine which is the most suitable feature detector to employ. Determining the scene content is a challenging task and it is even harder when the images to analyse have transformations, which can reduce dramatically the effectiveness of many of the methods currently available.

1.2 Contributions

The contributions made during the course of this research are summarised below.

- Predicting the performance of a feature detector operating in a complex and unknown environment is a challenging task. As mentioned above, many of the performance metrics and indicators available not always reflect the actual performance of a feature detector or they are reliable only within a specific range of applications. This thesis work proposes an evaluation framework based on the improved repeatability proposed in [6]. This method allows to determine how a detector will perform under varying transformation type and amount. In order to obtain reliable results, the proposed framework has been designed assuming the availability of a large image database including a wide variety of scenes well representing real-word scenarios.

- Feature detectors are heavily influenced by the operating context and a universal feature detector which is suitable for any application is not available yet. From the perspective of a vision application developer, a possible design solution is to run multiple feature detectors at the same time and then utilise the best feature set obtained. The most relevant drawback of this approach is the demand of computational resource required, which increases as the number of detectors employed. In this thesis is proposed an alternative approach consisting of selecting the optimal feature detector, where a detector is considered the optimal for a particular operating condition when it is expected to achieve the best possible

performance than any others available for the selection. A working prototype has been developed and assessed against several state-of-the-art feature detectors.

- A comprehensive performance prediction model of a feature detector should take into consideration both the scene content and image changes. Thus, a second evaluation framework to investigate how those factors influence detectors' performance is presented in this thesis. A new metric, trait indices, is introduced to measure the tendency of feature detectors to obtain their best and worst performance with scenes having particular characteristics (e.g. outdoor or indoor) or containing particular types of elements (e.g. human-made or natural). Several feature detectors have been assessed with this new framework and the results obtained are presented and discussed in this thesis.

1.3 Thesis Structure

An outline of the chapters of this thesis is as follows.

Chapter 2: Introduces local invariant feature detection principles and presents a survey of the most relevant methods available in literature such as SIFT, MSER and Harris-Affine detectors, which are also utilised for the experimental results presented in the chapters from 3 to 5. The chapter ends with a section dedicated to an overview of the most comprehensive and relevant evaluation methods and performance metrics utilised to evaluate local feature detectors.

Chapter 3: A new evaluation framework is introduced and utilised to characterise the performance of several state-of-the-art local feature detectors under varying JPEG compression, Gaussian blur and light reduction changes. Utilising the improved repeatability [6], the evaluation framework provides a punctual measure of the performance for each transformation type and amount available in the image database [7] and allows to draw some conclusions about the influence of the scene type and content on detectors' performance.

Chapter 4: A tool capable of selecting the most suitable feature detector in relation to the transformation of a target image is presented in this chapter. The tool employs a selection criterion based on the performance characterization of several detectors obtained with the evaluation framework from Chapter 3. Thus, the proposed tool can select the optimal feature detector for any given type and amount of JPEG compression, Gaussian blur and light changes available in the image database [7]. The results prove that the proposed tool has a good accuracy and requires a short execution time to determine the optimal feature detector.

Chapter 5: The limit of a model considering only image changes is discussed and an evaluation framework to characterise local feature detectors' performance in relation to the scene content is proposed in this chapter. Trait indices are introduced as a metric to quantitatively express how a detector performs with a particular type of scene. The set of local feature detectors already assessed in Chapter 3 are examined under varying JPEG compression, Gaussian blur and light changes from the perspective of the scene content, hence their traits indices are computed and discussed in order to draw conclusions on how the scene content influences a detector's performance.

Chapter 6: Provides a summary of the work presented in this thesis and draws important conclusions. The future directions of this research are outlined and discussed.

2

Local Invariant Feature Detectors and Their Evaluation

This chapter, after a short introduction to the problem of feature detection and image matching, provides a survey on the most relevant feature detectors available in literature and on the methods to evaluate their performance.

2.1 An Introduction to Feature Detection

A local feature is a relevant piece of information in an image, which is generally related to the variation of one or more properties and then it is well distinguishable by its immediate neighbourhood [1]. Corners, edges and blobs are examples of local features. Once a feature is detected, the surrounding region is utilised to compute a descriptor, which is used for image matching. One of the most popular and robust descriptor is SIFT [8], however there are many other good alternatives that are frequently employed in visual applications such as SURF [9], GLOH [10], HoG [11] and ORB [12].

In real-world applications, images can have different types of geometric and photometric transformations such as scaling and illumination changes. In order to succeed in image matching tasks, a local feature detector must be invariant to such image transformations, so that the same features are correctly detected in different images representing the same scene. For example, in the case of viewpoint change, the corners and edges must be detected in both the images of the same scene, even if they appear different due the different observation point as shown in Figure 2.1, where several features are matched in two images from the Graffiti dataset [13]. In the context of image matching, the repeatability of the features detected in different images of the same scene is one of the most desirable properties for a feature detector [15]. It is given as the ratio

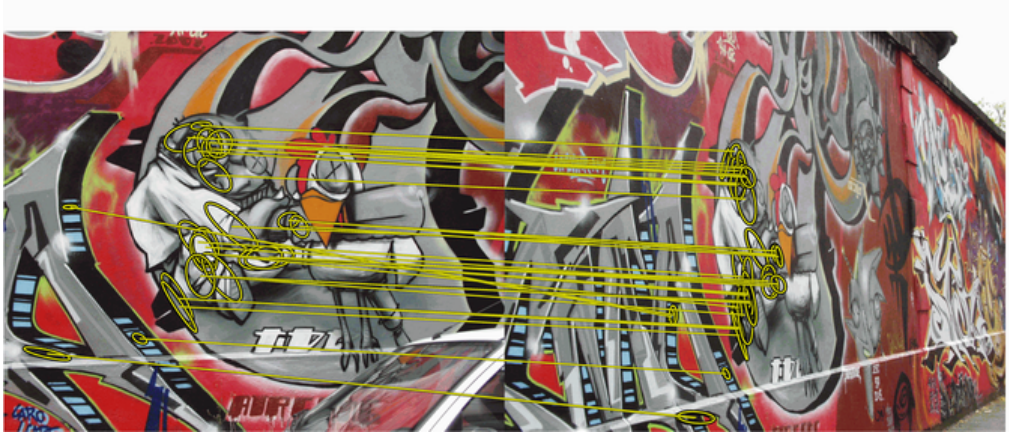

Figure 2.1: Feature matching example [14].

between the repeated features and the repeatable features between two images, target and reference, and it can be interpreted as the resilience of a feature detector to image transformations. However, repeatability is not the only property desirable for a feature detector. In [3] some of the most important characteristics of local invariant feature detectors are described. The localization accuracy is very important for several applications as 3D object reconstruction [1]; a sufficient number of interest points are required by certain applications such as object and scene recognition [1]; the coverage of the extracted set of features is an important property for some applications as grid calibration and homography estimation [16]; furthermore, a feature detector should detect features efficiently in order to be suitable for real-time applications or, more in general, where the execution time is a critical constraint.

In the following sections, a selection of state-of-the-art local feature detectors is described followed by an overview of the most relevant methods and metrics to evaluate their performance.

2.2 State-of-the-art Local Invariant Feature Detectors

In this section are discussed a selection of the most commonly used local invariant feature detectors in modern vision applications. The methods discussed are: Scale Invariant Feature Transform (SIFT) [8], Speeded-Up Robust Feature (SURF) [9], Harris-Laplace (HARLAP), Harris-Affine (HARAFF), Hessian-Laplace (HESLAP), Hessian-Affine (HESAFF) [2], Edge-Based Region (EBR) [17], Intensity-Based Region (IBR) [18], Maximally Stable External Region (MSER) [19], Salient Regions (SALIENT) [20], and Scale-invariant Feature Operator (SFOP) [21].

2.2.1 Scale Invariant Feature Transform (SIFT)

The Scale Invariant Feature Transform (SIFT) [8] algorithm integrates an efficient local feature detector and a highly distinctive descriptor. The feature detection stage includes two steps. The first one consists in searching all scales and image locations to identify the potential interest points. Then, a 3D quadratic interpolation technique is employed

to localise them with sub-scale and sub-pixel accuracy. The SIFT method approximates Laplacian-of-Gaussian (LoG) with a Difference-of-Gaussian (DoG) to extract in a time efficient manner blob-type features, which are assigned with an orientation depending upon the local image gradient direction.

The last stage of SIFT computes a descriptor consisting of 128 coefficients, based on the histogram of local oriented gradients around each of the interest points. The SIFT's descriptor is considered robust and particularly suitable for object recognition tasks [1]. However, the large amount of computation required by the descriptor makes SIFT not suitable for real-time applications [9].

2.2.2 Speeded Up Robust Feature (SURF)

SURF [9] is a scale invariant algorithm that includes three stages: feature detection, feature description and feature matching. Initially, the algorithm operates at various scales convolving rectangular masks of increasing size with the input image in its integral representation. By means of a sampling process on the resulting series of blob response maps, a set of candidate features is extracted and then filtered by a threshold to select the high-contrast features. The descriptor is based on a sum of Haar wavelets and it is computed for each selected feature after orientation assignment to render SIFT rotation invariant. The last stage matches the image features on the basis of local descriptors by applying neighbour matching scheme [8].

2.2.3 Harris-Laplace and Harris-Affine

The Harris-Laplace (HARLAP) [2] is a scale invariant detector consisting of two stages: a multi-scale Harris corner detector is utilised to determine the location of the image features followed by a Laplacian operator that is responsible for selecting the scale of local structures.

The Harris-Affine (HARAFF) [2] detector is built on top of the Harris-Laplace detector. Iteratively, for each of the features obtained by the Harris-Laplace detector, the algorithm determines the affine regions around the interest points with the second-moment

matrix. The resulting affine regions are then normalised and approximated with circular areas and the new positions and scales in the normalised image are determined.

2.2.4 Hessian-Laplace and Hessian-Affine

The Hessian-Laplacian (HESLAP) [2] detector is based on the same principle utilised by the Harris-Laplace detector with the difference that the interesting point locations are determined by means of the determinant of the Hessian Matrix. As the Harris-Laplace detector, the Hessian-Laplacian utilises the Laplacian operator to determine the scale at which there is maximum similarity between the feature detection operator and the local image structures.
The Hessian-Affine (HESAFF) [2] detector estimates the affine regions around the local features extracted by the Hessian-Laplacian detector following the same process of the Harris-Affine detector, which is summarised in the previous section.

2.2.5 Edge-Based Region (EBR)

In an image, edges are stable features that can be detected over a range of photometric and affine transformations. Edge-Based Region (EBR) [17] algorithm includes several steps: Harris corners detector [22] finds a set of corners in the image; Canny's edges are determined [23]; a one-dimensional family of parallelograms is built; the resulting EBR's features set is a selection of those parallelograms, which is determined using local extrema of invariant function as a criterion. EBR, which can be classified as a corner-based feature detector, is considered to perform well on scenes characterised by the presence of structures and regions delimited by sharp edges.

2.2.6 Intensity-Based Region (IBR)

The Intensity-Based Region (IBR) [18] detector is often classified as a segmented-based detector. The IBR's approach is based on the detection of invariant blob-like structures in an image. The algorithm first selects the intensity extrema at multiple scales, then explores the surrounding area along radial directions. The interconnection of the maxi-

mum of the invariant function along each of the rays delimits an irregular region, which is replaced by the best fitting ellipse.

2.2.7 Maximally Stable Extremal Regions (MSER)

The features by the MSER [19], the Maximally Stable Extremal Regions, are selected among the resulting regions of a watershed-like segmentation, which are selected if stable over a range of thresholds. Those regions are often blob-like structure similar to the features detected by IBR, so MSER is often classified as a blob detector. Due to the use of a watershed-like segmentation, MSER is computationally efficient and considered to perform well on structured images characterised by uniform regions separated by strong intensity changes.

2.2.8 Salient Regions (SALIENT)

The Salient Region detector (SALIENT) [20] is based on information theory, in particular, it utilises the entropy for localising the so-called salient regions, which are characterised by high complexity or unpredictability. The entropy of the probability function centred on each pixel is evaluated and the set of entropy maxima over scale are computed in order to determine the candidate image features. The candidate regions are then ranked over the entire image using their saliency and a specific number of top ranked regions are finally selected as salient regions.

2.2.9 Scale-invariant Feature Operator (SFOP)

Scale-invariant Feature Operator (SFOP) [21] can be classified as a spiral detector. It is a scale-space extension of the feature operator proposed in [24] with the spiral model introduced by Bigűn *et al.* in [25]. SFOP aims to identify different type of complementary image features: corners, circles and blobs. The high complementarity of the detected features makes SFOP a good feature detector for object recognition tasks and camera calibration, particularly with poorly textured scenes [21].

2.3 Evaluating Local Feature detectors

Since the early pioneering works on local features appeared in the mid '50s, when visual researchers observed that the most relevant information of a shape is concentrated in a few dominant points [26], a number of different methods to detect such relevant points have been proposed. The introduction of SIFT in [27] and [8] stimulated this research field even further yielding to the availability of wide variety of methods to detect features in images. Consequently, the issue of evaluating them has assumed a relevant importance in vision research. The purpose of this section is to present a selection of the most representative contributions in this research field.

The localisation accuracy is a desirable property of a feature detector as it is very important for some visual applications such as 3D reconstruction and image tracking, thus it is frequently assessed in evaluation works. In [28] the localisation error is utilised as a performance indicator to compare several state-of-the-art invariant feature detectors such as MSER, HARAFF, HESAFF and SIFT. In [29] feature detectors are assessed in the context of the automatic image orientation systems. Localisation error is also an important component of the performance metric presented in [30] and [31] where various contour-based and local affine invariant detectors are compared to each other.

Repeatability is the ratio of the repeated features between a reference image and a target image with respect to the total (repeatable) features and it can be interpreted as the resilience of a feature detector to image transformations. Repeatability and information content are utilised in [15] to compare detectors under various geometric and photometric image changes such as viewpoint and illumination variations. The repeatability is also used in [32] to evaluate feature detectors in the context of image retrieval. A refined definition of repeatability is given in [2] and then used to compare state-of-the-art affine and scale invariant feature detectors with the HARAFF and HARLAP detectors, which were also introduced in [2]. The same repeatability definition is employed in [3]. Here, a set of six state-of-the-art invariant local feature detectors, representing a wide variety of different approaches, are compared utilising the Oxford's datasets [13], which includes

eight scenes and a six image transformations at various amounts. A repeatability-based metric, namely average repeatability, is proposed in [33] and utilised to compare corner detectors in [31] and [34]. In [6] is proposed an improved repeatability definition, which is proven to be consistent with the actual performance of a wide variety of feature detectors across well-established datasets.

The feature coverage and distribution have an effect on visual applications, such as homography estimation [16]. In [35] and [36] the coverage and complementarity of the detected features are utilised to evaluate and compare feature detectors. The completeness and complementarity of local feature detectors are investigated in [37] from the perspective of the information content representation of images. The idea investigated in [37] is that both the number and the distribution across an image of the extracted features is important, as they should cover all the relevant part of an image in order to allow a proper representation of the information contained in it.

Several other works employ different evaluation criteria and metrics. In [38] the performance of detectors is evaluated under viewpoint, scale and light changes by using a large database of images with recall rate as a performance measure, which is defined as the ratio between potential point matches and the number of total interest points in the compared images. An evaluation method based on visual inspection is proposed in [39] for assessing edge detectors. Canny's criteria [23] namely good detection, good localization accuracy and low-responses multiplicity, is used for a theoretical assessment of edge detectors in [40].

A feature detector can be evaluated indirectly by examining the overall performance of a vision application employing that detector. In [41] several feature detectors are evaluated for their suitability for visual SLAM applications. The evaluation in [42] takes place in the context of face detection, whereas a variety of local feature detectors are compared utilising object recognition tasks in [43] and [44].

3

An Evaluation Framework Utilising a Large Number of Scenes

The first step towards a robust method to select the optimal feature detector in relation to the imaging conditions is having a model that reliably describes a detector's performance. Many of the evaluation works available utilise datasets with a relatively small number of scenes and the results obtained do not provide a comprehensive picture of the real feature detectors' capabilities. Indeed, calculating the repeatability rate on a very small number of different scenes can result in either overestimation or underestimation of the real detectors' performance [5] [4]. This chapter proposes a new evaluation framework designed for taking advantage of a large image database including a wide variety of real-world scenes. The results for several state-of-the-art local feature detectors under varying JPEG compression ratio, blur and light changes are presented and discussed.

3.1 Introduction

In the last few years this research topic has been very active and several remarkable works have been proposed [3] [36] [45] [46]. However, most of the evaluations done so far are limited to utilise some standard datasets, such as Oxford database, [13] with only one or two scenes for each type of considered image transformation. Since the performance of feature detectors is highly dependent on the scene content, it is critical to evaluate them utilising a large number of scenes under varying imaging conditions. So far, one of the largest number of scenes which has been used is only 60 [38]. Moreover, the scenes employed in [38] are not real-world scenes and are captured in a highly controlled environment. To bridge this gap, a new evaluation framework has been designed by taking advantage of large image databases in order to provide a reliable performance evaluation of feature detectors. The proposed framework has been utilised for evaluating the performance of several state-of-the-art local feature detectors. The results obtained are presented and discussed later in this chapter along with the large image database utilised, which includes images from 539 different real-world scenes (nearly 9 times the size used in [38]) for each considered transformation.

The remainder of this chapter is organised as follows. Section 3.2 describes the image database utilised for the evaluation of the feature detectors; in Section 3.3, the evaluation framework in introduced; the comparison results obtained for several local feature detectors are presented and discussed in Section 3.4; finally, Section 3.5 summarise the work presented here and draws conclusions.

3.2 Image Datasets

As indicated above, most of the image database available are not suitable for the proposed framework due to the small number the scenes included or to their limited variety. On the contrary, the image database proposed by Ehsan *et al.* [7] that is utilised here, contains a large number of images, 20482, from 539 real-world scenes representing a wide variety of outdoor and indoor environments and including both natural and human-

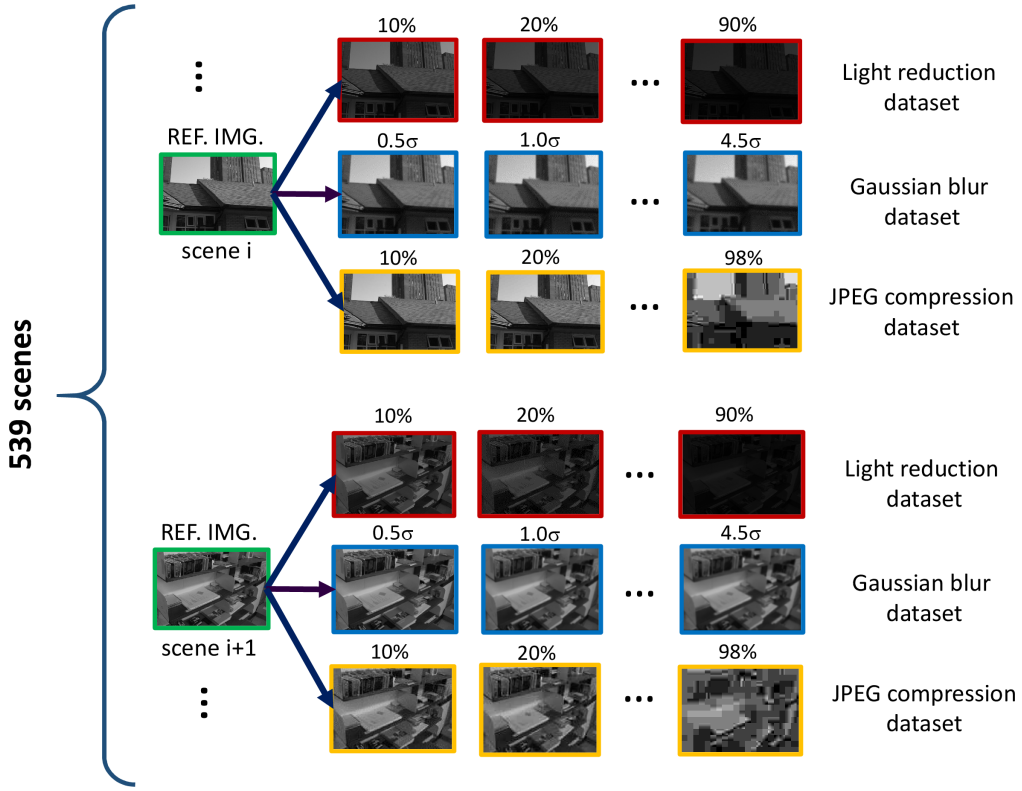

Figure 3.1: Each of 539 scenes included in the database has undergone increasing amounts of light reduction, blurring and JPEG compression to generate the datasets.

made elements. As shown in Figure 3.1, each of the 539 scenes has been undergone to increasing amounts of different transformations in order to obtain three datasets each for JPEG compression, uniform light and Gaussian blur changes. Each dataset includes the original camera shot as a reference image for a particular scene and several images of the same scene with different amounts of a transformation for a total of 10 images for Gaussian blur (1 reference + 9 transformed images) and 14 for JPEG compression and uniform light changes (1 reference image + 13 transformed images). JPEG compression rate is varied as follows for the 13 transformed images: 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, and 98%. Similarly, light brightness is reduced in 13 steps of 10%, 20%, 30%, 40%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, and 90%. The 9 blurred images are obtained with a Gaussian filter with increasing σ: 0.5, 1, 1.5, 2, 2.5, 3, 4 and 4.5. Figure 3.2 provides a sample of the scenes available in the database [7] and shows the related transformed images.

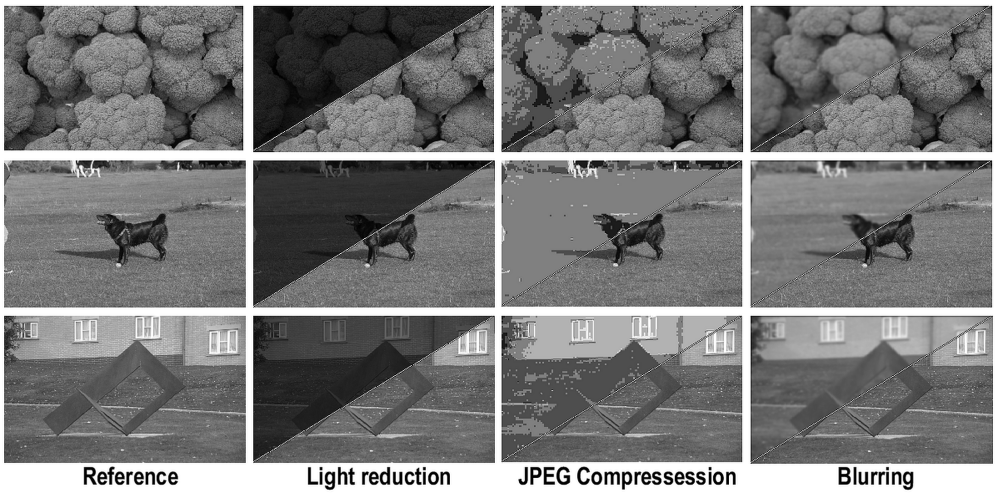

Figure 3.2: The reference image of three scenes and the effect of the application of 60% of light reduction, 98% of JPEG compression rate and 4.5 σ Gaussian blur.

3.3 Performance Evaluation Framework

The proposed framework is based on the improved repeatability measure presented in [6], as it is consistent with the actual performance of a wide variety of feature detectors across well-established datasets. The repeatability rate defined as follows:

$$Repeatability = \frac{N_{rep}}{N_{ref}} \tag{3.1}$$

where N_{rep} is the total number of repeated features and N_{ref} is the number of interest points in the common part of the reference image and the test image.

The proposed evaluation framework has been designed assuming the availability of and image database (I) including a group of (n) of datasets for each of the image transformation (t) considered for the analysis. Each of the datasets includes the original image of a scene (reference) and a series of transformed images (targets) obtained by applying a transformation at increasing amount. Thus, each dataset is generated from a single scene utilising a particular transformation as detailed in Section 3.2.

The proposed evaluation framework consists of the following steps, which are repeated for each image change considered for the feature detector comparison.

STEP1: Let D denotes the set of the assessed feature detectors, then choose a detector $d \in D$ and, for each of the images in each dataset in I, compute the repeatability score using the image with no transformation as a reference image. The amount of a specific image transformation (uniform light changes, blur variations and JPEG compression changes) is varied in m discrete steps. Let A be the set of indices representing such steps of increasing transformation amounts where '1' correspond to the reference image:

$$A = \{1, 2, 3, \ ... \ ... \ , m\} \tag{3.2}$$

Let B_{kd} be the set of repeatability rates computed for any one specific step $k \in A$ for the feature detector d in an image database containing n scenes:

$$B_{kd} = \{r_1, r_2, \ ... \ ... \ , r_n\} \tag{3.3}$$

A set B_{kd} includes a repeatability rate obtained from the k^{th} target image of each of datasets for the considered transformation which are 539 in [7]. While, k varies from 1 to m, where m is 10 for blur and 14 for JPEG compression and light changes for getting results shown later in this chapter (Section 3.4).

STEP2: For every feature detector, the arithmetic average of the repeatability scores is computed for each amount of image transformation. Let C_d be the average curve for the detector d:

$$C_d = \{\langle B_{1d}\rangle, \langle B_{2d}\rangle, \ ... \ ... \ , \langle B_{md}\rangle\} \tag{3.4}$$

$$\langle B_{kd}\rangle = \frac{1}{n}\sum_{i=1}^{n} r_i \quad where \quad r_i \in B_{kd} \tag{3.5}$$

STEP3: For every feature detector, the number of scenes for which it achieved the best repeatability among all considered detectors over any step of image transformation amount is computed. Let B_k be the set of all repeatability rates of all detectors d

corresponding to all the n images at the discrete step k of transformation amount:

$$B_k = \bigcup r_j \quad \forall r_j \in B_{kd}, \forall d \in D, \forall j \in \{1..n\} \tag{3.6}$$

Let $B_{k(n)}$ and $B_{k(n-1)}$ be the highest and the second-highest value in B_k. The share T_{kd} of scene images under the amount $k \in A$ of transformation for which the detector d scored the highest repeatability is given by the following equations:

$$T_{kd} = \frac{S_{kd}}{n}, \quad \text{where} \quad S_{kd} = \sum_{i=1}^{n} x_i \tag{3.7}$$

$$x_i = \begin{cases} 1/h & \text{if } r_i = B_{k(n)} \quad r_i \in B_{kd} \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

where S_{kd} is a value which represents the number of times the detector d scores the best repeatability at the step k; whilst the score function x_i assumes a value different from 0 only when the repeatability score for the i^{th} dataset is the highest. In Equation 3.8, h is the normalization coefficient; it is equal to the number of feature detectors whose repeatability is equal to $B_{k(n)}$ and has the function of keeping the shares consistent: $\sum_{d \in D} S_{kd}/n = 1$.

STEP4: For every feature detector, the number of scenes for which it scores the highest or the second-highest repeatability is computed. The score function becomes:

$$x_i = \begin{cases} 1 & \text{if } r_i \geq B_{k(n-1)} \quad r_i \in B_{kd} \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

STEP5: This step introduces a threshold in the score function. The detectors with a repeatability that is at least the 90% of the highest mark a point:

$$x_i = \begin{cases} 1 & \text{if } r_i \geq 0.9 \cdot B_{k(n)} \quad r_i \in B_{kd} \\ 0 & \text{otherwise} \end{cases} \tag{3.10}$$

This step, together with the previous one, has the purpose of showing if a detector can

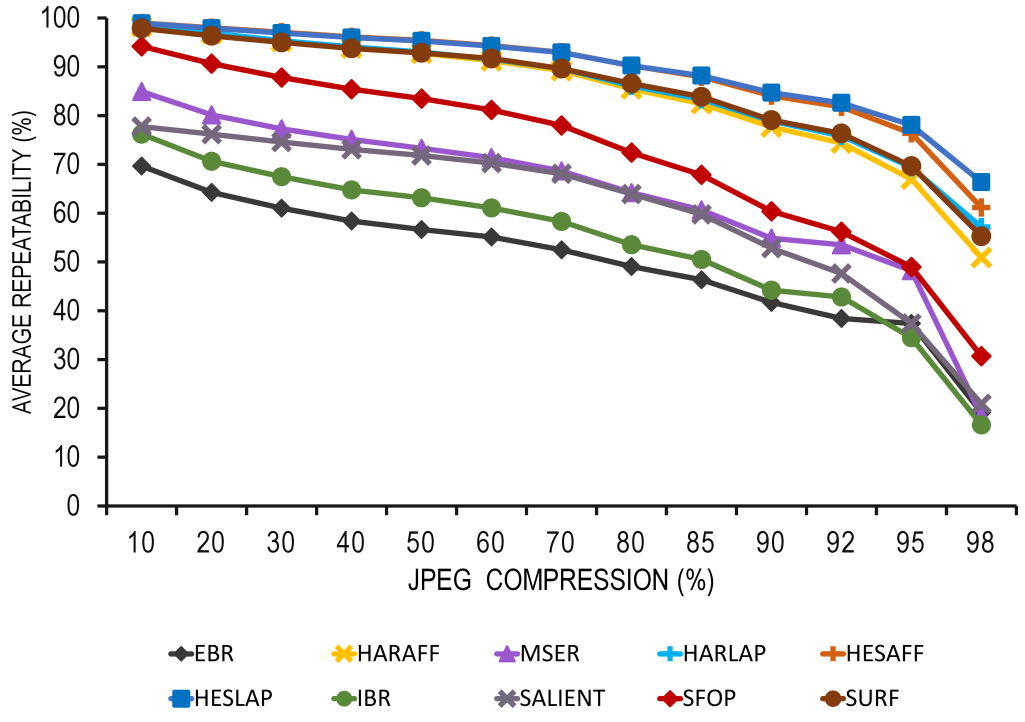

Figure 3.3: Average repeatability curves for JPEG compression.

be a proper replacement for another one. In particular, the threshold in Equation 3.10 relates the information from Equation 3.9 with repeatability performance. Indeed, once that detector is identified which scored the best for a particular scene, a lower ranked detector is considered a valid option only if its repeatability rate is within the threshold.

3.4 A Comparison of Local Feature Detectors

Utilising the proposed framework, a performance comparison of several state-of-the-art detectors under varying JPEG compression ratio, blur and uniform light changes is done with the large image database available at [7]. The selected detectors are representative of a wide variety of different approaches (Section 2.2) and include the followings: Edge-Based Region (EBR), Harris-Affine (HARAFF), Hessian-Affine (HESAFF), Harris-Laplace (HARLAP), Hessian-Laplace (HESLAP), Maximally Stable External Region (MSER), Intensity-Based Region (IBR), Salient Regions (SALIENT), Scale-invariant Feature Operator (SFOP) and Speeded Up Robust Feature (SURF). The re-

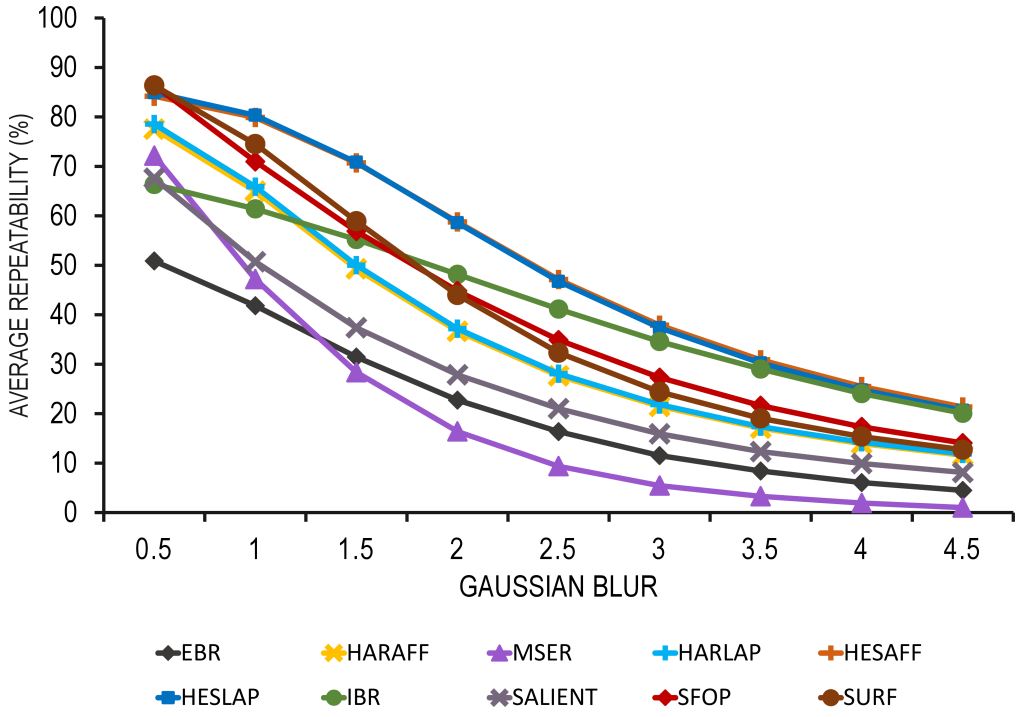

Figure 3.4: Average repeatability curves for Gaussian blur.

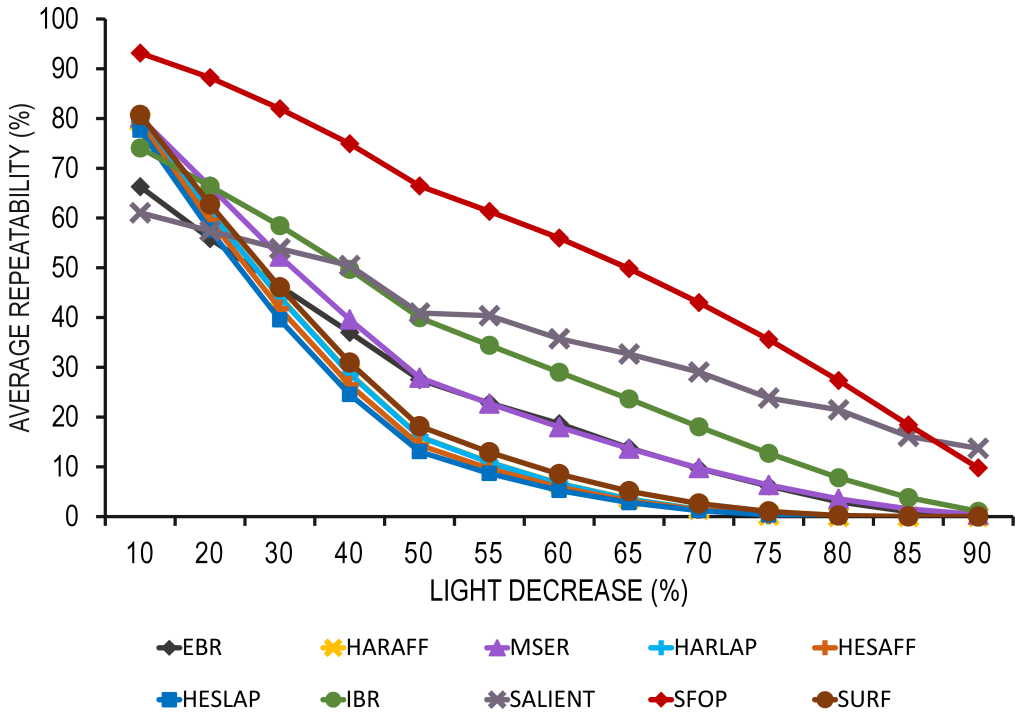

Figure 3.5: Average repeatability curves for uniform light reduction.

peatability data are obtained for each transformation type utilising the image database available at [7] and the authors' original programs with the control parameter values suggested by them. The feature detector parameters could be varied in order to obtain a similar number of extracted features for each detector. However, this has a negative impact on the repeatability of a detector [2] and is therefore not desirable for such an evaluation.

Figures 3.3 to 3.5 show the outcome of the average repeatability analysis. The average curves for JEPG compression (Figure 3.3) present a relatively small slope denoting a good resilience of the detectors against this transformation. HESLAP and HESAFF are the most suitable detectors for dealing with JPEG compression. Their performances are very close to each other with the average repeatability which decreases slowly from 98.9% to a value around 77% before dropping to 61.1% for HESAFF and to 66.4% for HESLAP at 98% of compression rate. From Figure 3.4 can be appreciated that HESLAP and HESAFF appear the most suitable detectors in the presence of Gaussian blur transformation. The IBR detector presents the highest stability against blurring with a slow decrease of the average repeatability from 80% at 0.5σ to 20% at 4.5σ. Although SFOP presents by far the best performance for light changes, the impact of this transformation is substantial on every detector (Figure 3.5). Even HESLAP and HESAFF, which captured the highest ranks in coping with JPEG compression and blurring transformations, are so much affected by light reduction that their average repeatability is roughly zero for 70% of light reduction and onwards.

The average repeatability curves presented in Figure 3.3 confirm the good performance of the Hessian and Harris-based detectors with JPEG compression as shown in [3], whereas the results for light reduction are very different. Indeed, the average repeatability of MSER is lower than the values recorded for IBR and SALIENT for light reduction amounts greater than 10%, whereas in [3] it is MSER that outperforms the other two detectors. In addition, from Figure 3.5, it can be inferred that feature detectors are very sensitive to light changes, which is not evident in [3].

Figures 3.6, 3.7 and 3.8 present the pies showing the percentage of scenes for which

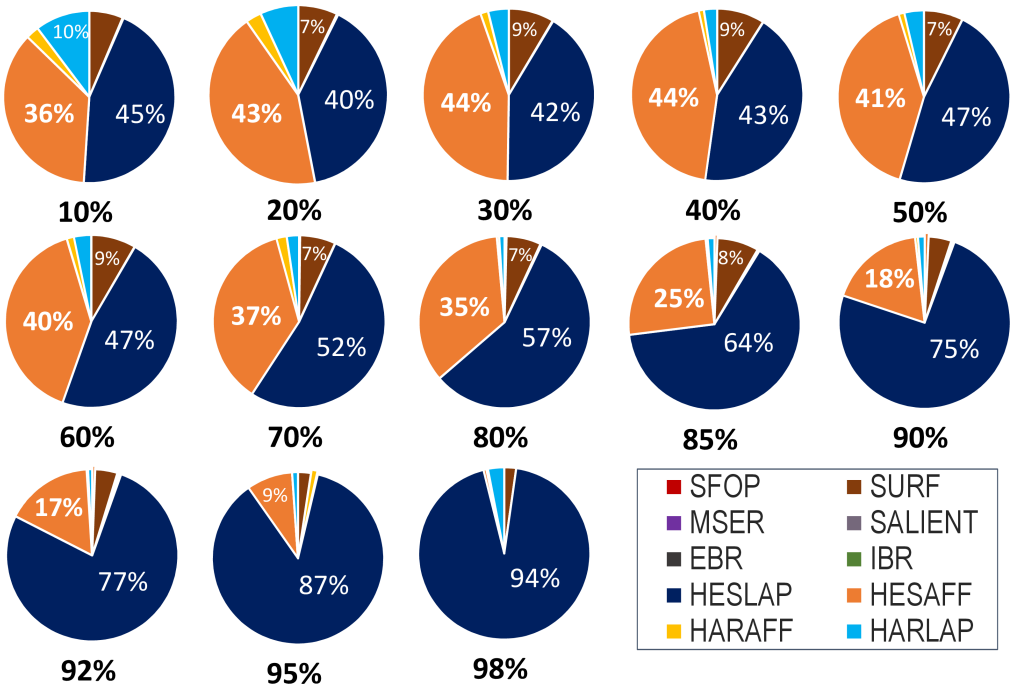

Figure 3.6: Ratio of the best repeatability for JPEG compression.

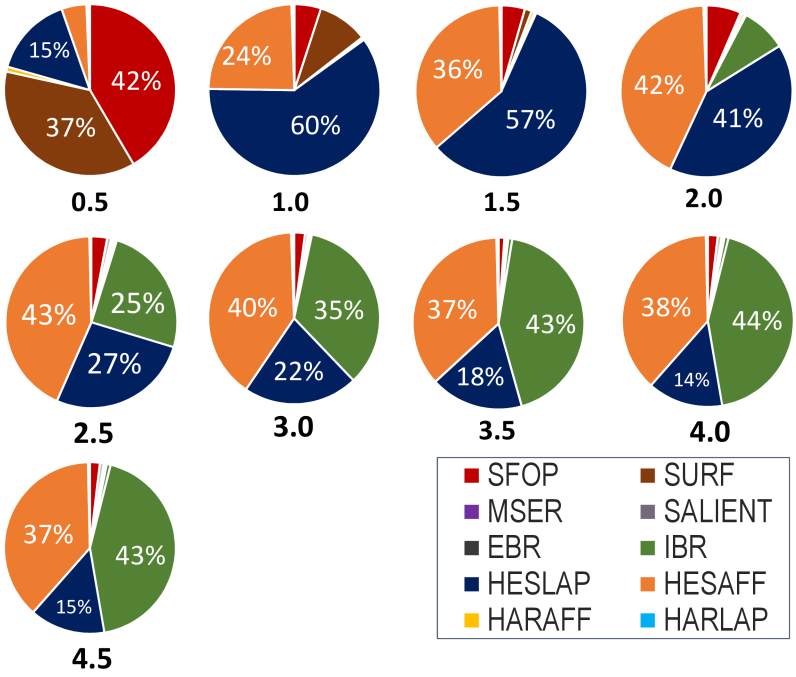

Figure 3.7: Ratio of the best repeatability for Gaussian blur.

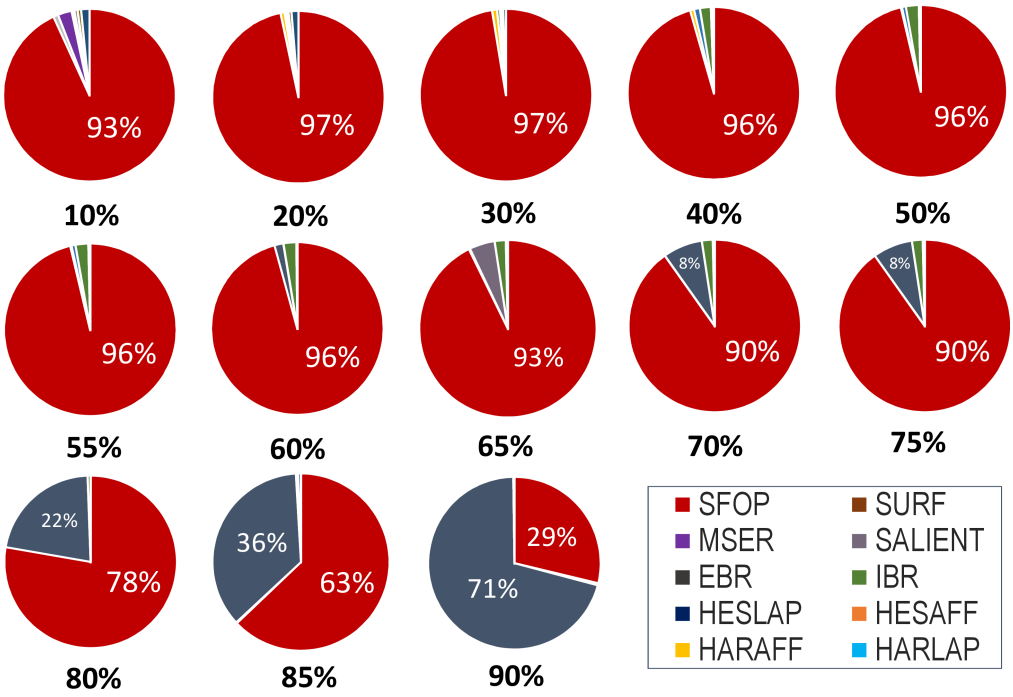

Figure 3.8: Ratio of the best repeatability for uniform light reduction.

each detector achieved the best repeatability among all others under JPEG compression, blurring and light changes respectively. The pies reveal that the local feature detectors are not invariant to the scene content. This is particularly evident for IBR under blurring (Figure 3.7) which presents comparable shares with HESAFF and HESLAP in spite of their average repeatability which is constantly higher than IBR's under JPEG compression (Figure 3.6).

Finally, the results obtained by applying the steps 4 and 5 confirm that HESAFF and HESLAP have good interchangeability for blurring and JPEG compression (Figure 3.9.a and Figure 3.9.b) as the application of the 90% threshold denotes that the repeatability rates of those two detectors are frequently close to each other. SFOP, instead, does not have any good alternative under light reduction. Indeed, the application of the threshold shows that the repeatability scored by SFOP is by far larger than the rate obtained by the second best detector for most of the scenes. This trend is kept for each amount of transformation except for highest (90%) where SALIENT achieves the best highest or second highest repeatability score ratio (Figure 3.9.c).

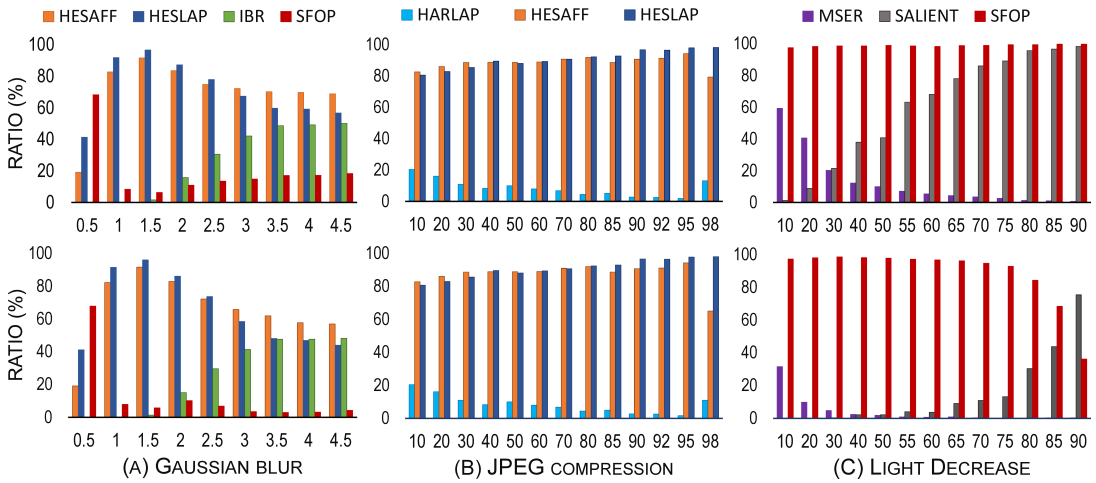

Figure 3.9: The highest or second-highest repeatability score (bottom) with and (top) without the 90% threshold for any amount of (a) Gaussian blur, (b) JPEG compression, and (c) light reduction.

3.5 Summary

The evaluation framework introduced in this chapter has been utilised to compare the performance of several local feature detectors under varying transformations. Although the results are presented only for JPEG compression, Gaussian blur and light reduction, the framework is applicable to any type of transformation and with any feature detector for which the improved repeatability rate is computable [6]. The results confirm that feature detectors are sensitive to the content of the scenes. Indeed, looking at the pies (Figures 3.6 - 3.8) it is straightforward to notice that several detectors achieve the highest repeatability on a significant share of the scenes even if their average repeatability is relatively low if compared to the highest values. This means that for some scenes the detector which usually perform the best, is outperformed by detectors which perform worse than it on average. A significant example of this behaviour is given by IBR under Gaussian blur whose average repeatability not always reflect the relatively wide pie shares obtained by this feature detector.

The relation between the scene content and feature detectors' performance is further investigated in Chapter 5 whilst the next chapter presents a tool for selecting the optimal

feature detector employing a selection criterion based on the average curves discussed in Section 3.4.

4

Automatic Selection of the Optimal Local Feature Detector

The results presented in the previous chapter confirm that a feature detector is suitable only for a limited range of imaging conditions. For example, SFOP performs better than any other detector under light reduction but it outperformed by SURF for JPEG compressed images. In this chapter is presented a first attempt to achieve a fully adaptive feature detection system that, given an input image, selects the feature detector which is expected to perform the best accordingly with the results presented in Chapter 3.

4.1 An Outline of the Proposed Approach

Local feature detection is an important and challenging task in most vision applications. A large number of different approaches have been proposed so far [1]. All these techniques present various strengths and weaknesses, which make detectors' performance dependent on the application and, more generally, on the operating conditions, such as the transformation type and amount [3] [5]. To overcome this problem, an obvious solution is to run multiple feature detectors so that the shortcomings of one detector are countered by the strengths of the other detectors. However, the computational demand of such an approach can be high and increases with the number of detectors employed. An alternative solution consists of a tool capable of automatically selecting the optimal feature detector to cope with any operating conditions as suggested in [1]. The work described in this chapter aims to bridge this gap by proposing a tool which can determine the transformation type (T) and amount (A) of input images and then select the detector that is expected to perform the best under those particular operating conditions. The proposed approach requires to have a prior knowledge of how feature detectors perform under any of the considered operating conditions (T, A). So, in order to design an effective selection stage (Figure 4.1), the evaluation framework introduced in Chapter 3 is utilised to characterise the performance of a set of feature detectors under varying transformation types and amounts. This performance characterisation, as well as the results presented later in this chapter, are obtained with the image database available at [7], which is described in details in Section 3.2. This image database includes 539 scenes, which has been used for generating the datasets for three transformations, namely light reduction, JPEG compression and Gaussian blur. Each dataset has a reference image and several target images, which are obtained by the application of the same transformation to a reference image with increasing amounts. Considering that the JPEG and light reduction datasets include 13 target images and a blur dataset has 9 target images, the resulting number of operating conditions available in the image database [7] is 18865.

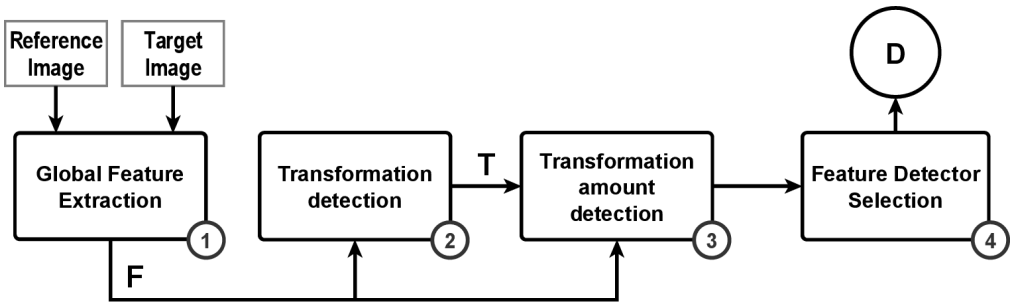

Figure 4.1: Block diagram of the automatic selection system; stage 1 extracts global features from input images, stages 2 and 3 determine the operation conditions, whereas stage 4 selects the optimal feature detector.

The remainder of this chapter is organised as follows. The proposed selection tool is introduced in Section 4.2 while the results of the tests are shown and discussed in Section 4.3. Finally, Section 4.4 draws important conclusions and discusses the next steps towards a more refined selection system.

4.2 The Automatic Selection Tool

The proposed system consists of four stages (Figure 4.1). The first stage extracts global features from the input images, then the second and the third stages determine the type (T) and the amount (A) of transformation respectively. The last one selects the optimal detector, which is expected to obtain the highest repeatability. The following subsections describe those four stages of the proposed system and provide more details about the selection criterion of the optimal feature detector.

4.2.1 Global Feature Extraction

The first stage analyses the input pair of target and reference images and then builds a vector of three features: $F = [f_L, f_B, f_J]$. The component f_L is the light reduction feature and is computed as the ratio between the mean of the image histogram of the target and the reference images: $f_L = h_t/h_r$. Hence, lower values of f_L correspond to higher amount of light reduction. The blur amount of an image is estimated with the perceptual blur metric proposed in [47]. The Gaussian Blur feature, f_B, is computed as

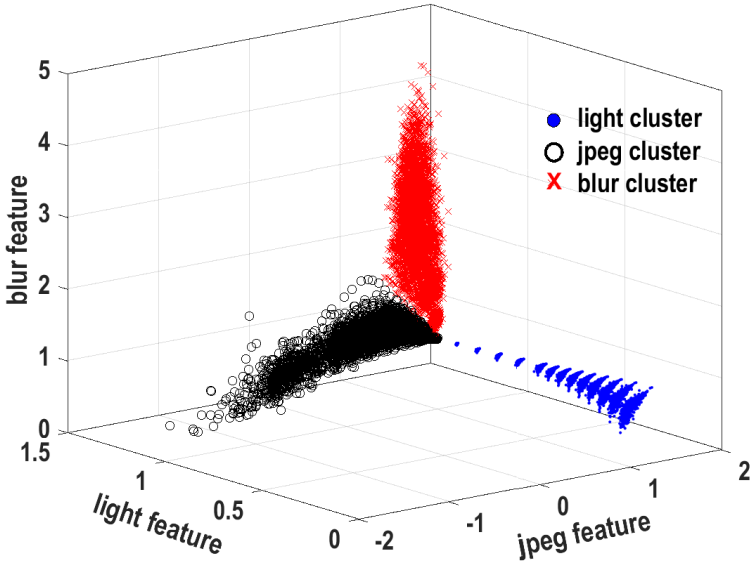

Figure 4.2: Set of features obtained from training set.

the ratio of the perceptual blur indices of the target and reference images respectively: $f_B = b_t/b_r$. A high value of f_B corresponds to a relatively high level of blurring in the target image. The JPEG feature f_J is computed with the reference-less quality metric proposed in [48], which produces a quality index of an image by combining the blockiness and the zero-crossing rate of the image differential signal along vertical and horizontal lines. Higher the compression rate of a JPEG image, lower is the value of f_J.

4.2.2 Transformation Type Detection Stage

The transformation (T) is determined with a Support Vector Machine (SVM) classifier with a linear kernel function. The SVM has been trained utilising a portion of the datasets [7] of 339 scenes chosen randomly. The related datasets for light changes, JPEG compression and Gaussian blur are employed to train the classifier. This results in a training set of 11865 feature vectors (13 x 339 for JPEG compression and light reduction, and 9 x 339 for blurring).

The features tend to form well-separated clusters as can be appreciated from Figure 4.2, which shows a plot of the feature vectors obtained with the test set. This allows

a high accuracy of the transformation type prediction, which is above 99%. Almost all the classification errors occur between blurred and JPEG compressed images at the lowest amounts of transformation (10-20% of JPEG compression rate and 0.5-1.0σ for Gaussian blur).

4.2.3 Transformation Amount Detection Stage

The third stage is composed of a set of SVMs, each specifically trained to predict the amount A of a single transformation type. So, once T is determined, the corresponding SVM is activated to determine the transformation amount from the feature vector F. The overall accuracy for light reduction is close to 100% while the percentage of transformation amounts correctly classified by the JPEG and blur SVMs are just 75% and 73% respectively. However, the results presented in Section 4.3, show the relatively low accuracy of the JPEG and blur classifiers do not significantly affect the overall performance of the automatic selection system.

4.2.4 Selection of the Optimal Feature Detector

This stage is implemented as a set of rules, which associate each pair (T, A) with the optimal feature detector D to operate under such type and amount of transformation. The evaluation framework from Chapter 3 is utilised to characterise the set of feature detectors available at runtime for selection. Such characterisation is carried out following the process described in Chapter 3 utilising the training set (Section 4.2.2) of 339 datasets per transformation. First, the improved repeatability rate [6] is computed for each feature detector using the authors' original programs and the parameters values suggested by them. The average of the repeatability rates is computed across all the scene images that are undergone to the same type and amount of transformation. For example, the average repeatability of a detector at 20% of JPEG compression is obtained as the mean of the repeatability scored with the 339 JPEG images compressed at 20%.

Utilising the outcomes of the performance characterisation, the optimal feature detector

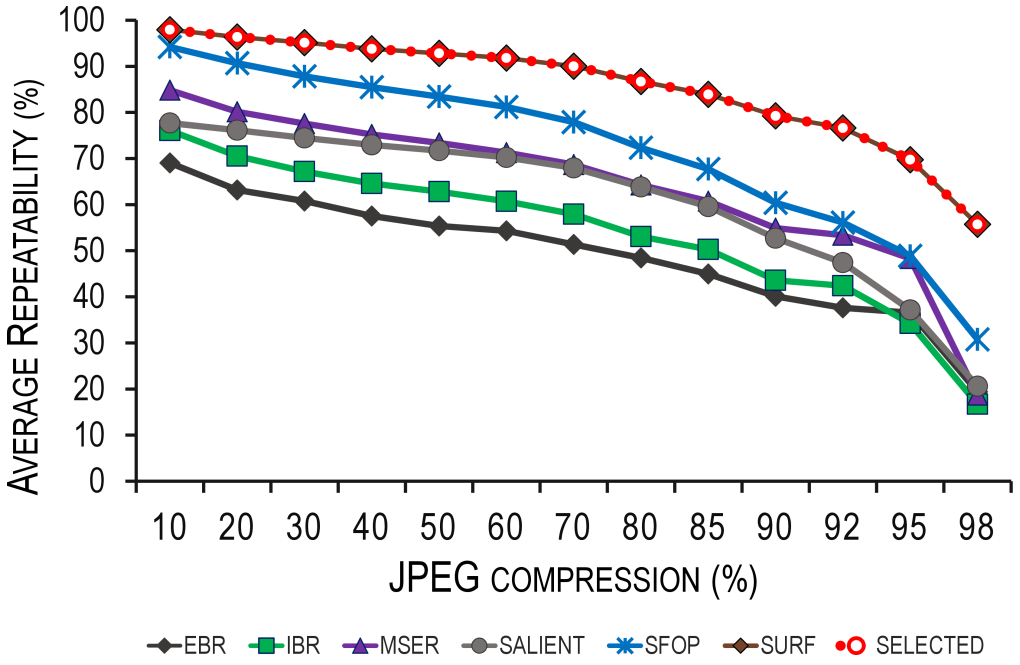

Figure 4.3: Average repeatability curves of the proposed selection tool and feature detectors working individually for JPEG compression.

for any operating condition is identified utilising the highest average repeatability as a criterion. The resulting set of associations, $(T, A) \rightarrow D$, is utilised by the proposed tool at runtime to select of the most suitable feature detector for any given input target image.

4.3 Test Results and Discussion

This section presents the results of the comparison between the selection algorithm and several feature detectors working individually under varying uniform light reduction, Gaussian blur and JPEG compression. The evaluation criteria are the accuracy, which is measured by means of the gap between the average repeatability of the best detector and the optimal detector selected by the tool, and the execution time. The employed set of feature detectors represents a variety of different approaches (Section 2.2) and includes the following: Edge-Based Region (EBR), Maximally Stable External Region (MSER), Intensity-Based Region (IBR), Salient Regions (SALIENT), Scale-invariant

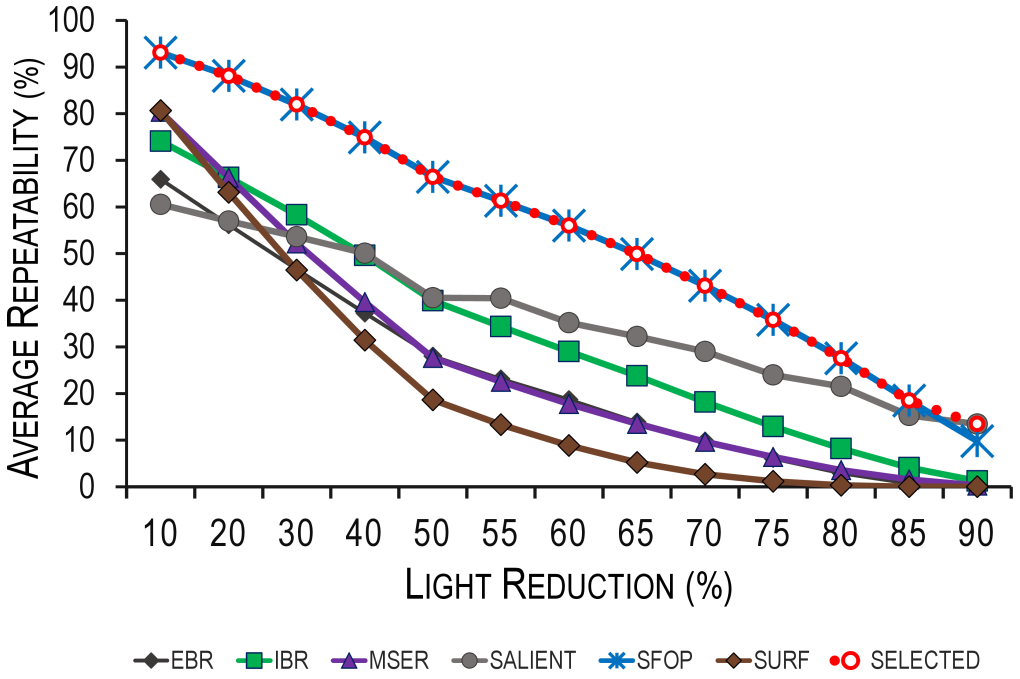

Figure 4.4: Average repeatability curves of the proposed selection tool and feature detectors working individually for light reduction.

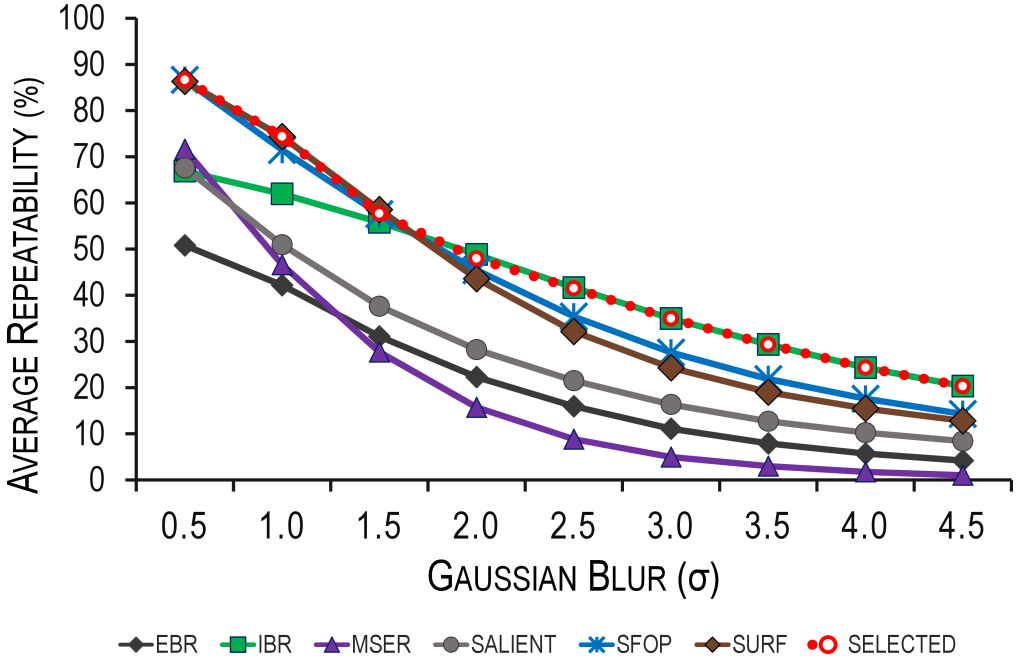

Figure 4.5: Average repeatability curves of the proposed selection tool and feature detectors working individually for Gaussian blur.

Feature Operator (SFOP), Speeded Up Robust Features (SURF). The scenes utilised for the tests are the remaining 200 scenes in [7], which are not included in the training set (Section 4.2.2). Thus, 200 datasets each for light reduction, JPEG compression and blurring transformations have been utilised as a test set. As it is done for characterisation of detectors' performance, the repeatability data are obtained using the original authors programs and with the recommended control parameter values suggested by them.

Figures from 4.3 to 4.5 show a comparison of the average repeatability of the feature detectors working individually and the selection algorithm (red dotted line) for the three transformations considered: JPEG compression, uniform light reduction and blurring respectively. Under JPEG compression, the accuracy of the selection is very high with a negligible gap error. Indeed, SURF performs the best under any transformation amount (Figure 4.3), so the accuracy of the selection depends only on the prediction of the transformation type, which is correct in more than 99% of the cases. The automatic selection tool performs well also with light reduction as it can be appreciated from Figure 4.4 where the red dotted line matches perfectly the SFOPs average curve up to 85% and the SALIENTs curve at 90% of light reduction.

To the contrary, under Gaussian blur, some selection errors occur as shown in Figure 4.5, where the gap between the average repeatability of the best detector and the one chosen as optimal by the selection tools is plotted. Between 1.5σ and 2.0σ (Figure 4.6) there is a dip of -1%. In that range of blurring intensity, the average curves of SURF and IBR intersect each other (Figure 4.5) and the wrong predictions of the transformation amount (A) cause some errors in the selection of the optimal feature detector. Although the probability that such classification error occurs is around 9%, the resulting gap error is just -1%. This is due to the little difference between the average repeatability values of SURF and IBR, which are close to each other at 1.5σ (58.54% vs 55.78%) and at 2.0σ (43.6% vs 48.8%).

A complete run of the proposed tool, from image loading to the detector selection, requires a time comparable to the fastest of the feature detectors considered: MSER. The

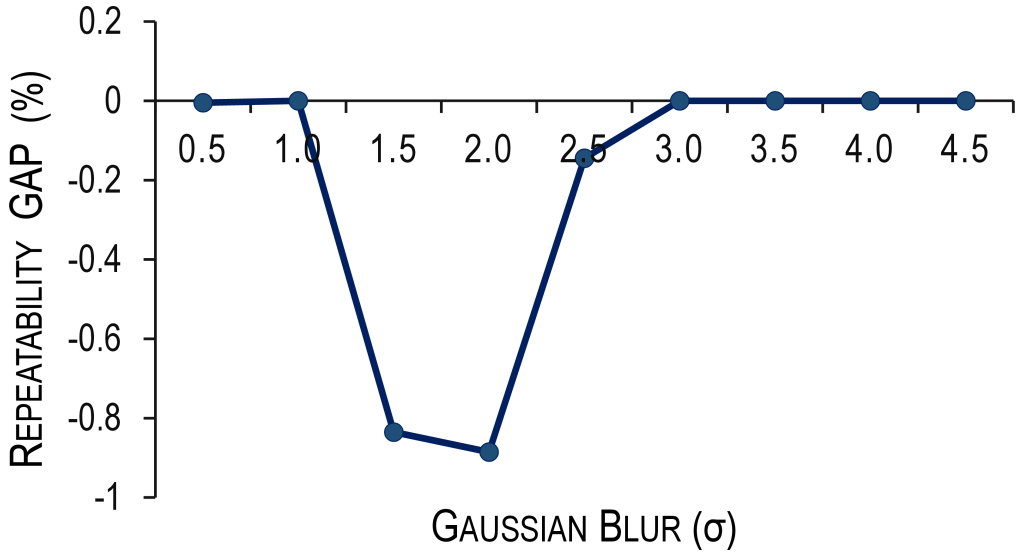

Figure 4.6: Average repeatability gap between the proposed selection tool and feature detectors working individually under Gaussian blur.

hardware employed for the test is a laptop equipped with a i7-4710MQ CPU, 16Gb of RAM, and a SATA III SSD Hard drive and the test images have a resolution of 1080 x 717 pixels. MSER and IBR are available as binary executables and have a running time of 150ms and 1.8 seconds respectively while the selection tool, which is a Matlab script, requires 170ms to load images and select a detector. Hence, a system employing the proposed tool with those two feature detectors can extract features in 170 + 150 ms (when MSER is optimal) or 170 ms + 1.8 seconds (when IBR is optimal) while running both MSER and IBR with an image and select the best, would require always more than 1.9 seconds. Thus, the proposed system is equally or more efficient than running more feature detectors with the same image, in addition, it scales really well with the number of feature detectors employed.

4.4 How Can the Selection Criterion be Refined?

The automatic tool for selecting the optimal feature detector proposed in this chapter represents an attempt to achieve a fully adaptive feature detector system capable of coping with any operating condition. The proposed approach is based on the knowl-

edge of the behaviour of detectors under different operating conditions, which are the transformation type T and the amount of such transformation, A. The next step towards a more robust automatic selection system is to consider the scene content as a part of the operating conditions. Indeed, a detector's performance depends also on that factor as confirmed by the comparison results presented and discussed in Sections 3.4 and 3.5. To bridge this gap, the next chapter introduces a new evaluation framework to investigate the influence of both image transformation and scene content on detectors' performance.

5

Performance Characterization in Relation to the Scene Content

The evaluation framework presented in Chapter 3 estimates detectors' performance by mean of reliable statistical indicators such as the average repeatability and pies. Although this approach allows inferring a correlation between repeatability and scene content, such a relationship cannot be explicitly and formally described with the method proposed. The aim of the evaluation framework proposed in this chapter is providing new insights into the behaviour of local feature detectors in relation to both the scene content and image transformation. The comparison results of the same set of local feature detectors assessed in Chapter 3 are presented later in this chapter, after the introduction of this new evaluation framework.

5.1 Introduction

Although the literature offers a variety of comparison works focusing on performance evaluation of image feature detectors under several types of image transformations, the influence of the scene content on the performance of local feature detectors has received little attention so far. Indeed, most of them focus mainly on characterising feature detectors' performance under different image transformations without analysing the effects of the scene content in detail. In [49], the feature tracking capabilities of some corner detectors are assessed utilising static image sequences of a few different scenes. Although the results permit to infer a dependency of the detectors' performance on the scene content, the methodology followed is not specifically intended to highlight and formalise such a relationship, as no classification is assigned to the scenes. The comparison work in [3] gives a formal definition for textured and structured scenes and shows the repeatability rates of six feature detectors. The results provided by [3] show that the content of the scenes influences the repeatability but the framework utilised and the small number of scenes included in the datasets [13] do not provide a comprehensive insight into the behaviour of the feature detectors with different types of scenes. In [50], the scenes are classified by the complexity of their 3D structures in complex and planar categories. The repeatability results reveal how detectors perform for those two categories. The limit in the generality of the analysis done in [50] is due to the small number and variety of the scenes employed, whose content are mostly human-made. The main goal of this work is to identify the biases of these detectors towards particular types of scenes, and how those biases are affected by three different types and amounts of transformations (JPEG compression, blur and uniform light changes). The methodology proposed utilises the improved repeatability criterion in [6] and the large image database detailed in Section 3.2 and consisting of 539 different real-world scenes containing a wide variety of different elements.

The remainder of the chapter is organised as follows. In Section 5.2, the proposed evaluation framework is described in detail. The results utilising the proposed framework

with several feature detectors are presented and discussed in Section 5.3. Section 5.4 provides a summary of the work presented in this chapter and the Section 5.5 discusses the limits of the proposed evaluation method.

5.2 The Proposed Evaluation Framework

The proposed framework has been designed by keeping in mind the objective of evaluating the influence of scene content on the performance of feature detectors. This framework has been designed to be utilised with the image database [7]. As explained in Section 3.2, that large image database (I) is organised in a series of n datasets. Each dataset contains images from a single scene with different amounts of an image transformation. The 539 scenes included in [7] are taken from a large variety of different real-world scenarios. The proposed framework consists of the steps discussed below.

5.2.1 Repeatability value sets

This framework is based on the repeatability criterion described in [6], whose consistency with the actual performance of a wide variety of feature detectors has been proven across well-established datasets [13]. As proposed in [6], the repeatability rate is defined as follows:

$$Repeatability = \frac{N_{rep}}{N_{ref}} \tag{5.1}$$

where N_{rep} is the total number of repeated features and N_{ref} is the number of interest points in the common part of the reference image.

Let D the set of the feature detectors to assess; then select one of them, $d \in D$, and for each of the images in each dataset of I, the repeatability scores are computed using the image with no transformation as a reference image while the amount of a specific image transformation is varied in m discrete steps. Let A be the set of indices representing such steps of increasing transformation amounts and P the set of indices representing the scenes included in I:

$$A = \{1, 2, 3, ...m\} \tag{5.2}$$

$$P = \{1, 2, 3, ...n\} \tag{5.3}$$

The value m corresponds to the maximum amount of transformation, 1 relates to the reference image (no transformation) and n is the total number of scenes in I. In particular, for the database at [7] n is 539 while m is equal to 10 for blurring and it is equal to 14 for JPEG compression and light changes.
The repeatability rates of a detector d are organized in m sets, each for a specific amount k of transformation:

$$B_{kd} = \{B_{1kd}, B_{2kd}, \ ... \ ...B_{nkd}\}, \ k \in A, d \in D \tag{5.4}$$

Each set B_{kd} includes n repeatability rates, one for each scene in I, which are 539 in [7].

5.2.2 Scene rankings

The top and lowest rankings for each detector d are built selecting the j highest and lowest repeatability scores at k amount of image transformation. Let $T_{kd}(j)$ and $W_{kd}(j)$ the sets containing the indices of the scenes whose repeatability falls in the top and lowest ranking respectively:

$$T_{kd}(j) = \left\{S_{kd(1)}, S_{kd(2)}, ...S_{kd(j)}\right\} \tag{5.5}$$

$$W_{kd}(j) = \left\{S_{kd(n)}, S_{kd(n-1)}, ...S_{kd(n-j+1)}\right\} \tag{5.6}$$

where $S_{kd(i)} \in P$ is the scene index corresponding to the i^{th} highest repeatability score obtained by the detector d for the scene under amount of transformation k. Thus, in accordance with this notation, $S_{kd(1)}$ is the scene for which the detector scored the best repeatability score, $S_{kd(2)}$ corresponds to the second highest repeatability rate, $S_{kd(3)}$ to the third highest and so on, until $S_{kd(n)}$ which is for the lowest one.

Figure 5.1: Some images from the database in the form by category

5.2.3 Scene classification

The scenes are attributed with three labels on the basis of human judgment. As described in Table 5.1, each label is dedicated to a particular property of the scene and has been assigned independently from the others. These attributes are: the location type (f), which may take the label outdoor or indoor, the type of the elements contained (g), which may take the label natural or human-made, and the perceived complexity of the scene (h), which may take the label simple or complex. Figure 5.1 shows a sample of the scenes from the image database [7] utilised for the experiments grouped so that each row shows scenes sharing the same value for one of the three labels f, g and h. For example, scene 9 is tagged as outdoor along with the scenes 128 and 380. At the same time, the first two are also classified as natural scenes whereas scene 380 includes mostly human made objects. Scenes 373, 40 and 295 are labelled as human-made and the first two is also classified as indoor. Scene 530 is categorised as a simple scene as it includes a few edges delimiting well-contrasted areas. Scene 76 is considered as natural, non-outdoor and complex, due to the rough surface of the broccolis that is information

rich.

Location Type	Outdoor	Indoor scene and close-up of a single or a few objects.
	Indoor	The complement of above.
Object Type	Human-made	Elements are mostly artificial.
	Natural	Elements are mostly natural.
Complexity	Simple	A few edges with quite regular shapes.
	Complex	A large number of edges with fractal-like shapes.

Table 5.1: Classification labels and criteria

5.2.4 Ranking trait indices

The labels of the scenes included in the rankings (Equations 5.5 and 5.6) are examined in order to determine the dominant types of scenes. For each ranking $T_{kd}(j)$ and $W_{kd}(j)$, the ratios of scenes classified as *outdoor*, *human-made* and *simple* are computed. Thus, three ratios are associated with each ranking where higher values mean higher share of the scene type associated:

$$\forall S_i \in T_{kd} : [F, G, H]_{T_{kd}} = \frac{\sum_i [f, g, h]_{S_i}}{j} \tag{5.7}$$

$$\forall S_i \in W_{kd} : [F, G, H]_{W_{kd}} = \frac{\sum_i [f, g, h]_{S_i}}{j} \tag{5.8}$$

These vectors contain three measures which represent the extent of the bias of a detector. For example, if the top ranking vector $[F, G, H]_{T_{kd}}$ presents $F = 0.1$, $G = 0.25$ and $H = 0.8$, it can be concluded that the detector d, for the given amount of image transformation k, works better with scenes where its elements are mostly natural (low G), with simple edges (high H) and that are not outdoor (low F). As opposed to that, if the same indices were for the lowest ranking it could be concluded that the detector obtains its lowest results for non-outdoor (F) and natural (G) scenes with low edge complexity (H).

5.3 Results

The proposed framework has been applied for producing the top and lowest rankings for a set of eleven feature detectors which are representative of a wide variety of different approaches (Section 2.2) and includes the following: Edge-Based Region (EBR), Harris-Affine (HARAFF), Hessian-Affine (HESAFF), Harris-Laplace (HARLAP), Hessian-Laplace (HESLAP), Scale Invariant Feature Transform (SIFT), Intensity-Based Region (IBR), Maximally Stable External Region (MSER), Salient Regions (SALIENT), Scale-invariant Feature Operator (SFOP), and Speeded Up Robust Feature (SURF). The remainder of this section is organised in two subsections: the first one provides details on how the repeatability data have been obtained and the second one is dedicated to the discussion about the trait indices of each local feature detector.

5.3.1 Repeatability Data

The repeatability data are obtained for each transformation type utilizing the image database [7] described in Section 3.2. These data are collected using the authors' original programs with control parameter values suggested by the respective authors. The feature detector parameters could be varied in order to obtain a similar number of extracted features for each detector. However, this has a negative impact on the repeatability of a detector [2] and is therefore not desirable for such an evaluation.

With the image database [7], 18865 repeatability rates have been computed for each local feature detector with the exception of SIFT, which has been assessed only under JPEG compression. It should be noted that SIFT detects more than 20,000 features for some images in database which makes it very time-consuming to do such a detailed analysis for SIFT. In the case of JPEG image database, it took more than two weeks to obtain results on HP ProLiant DL380 G7 system with Intel Xeon 5600 series processors. Therefore, results for SIFT are not provided here.

The number of datasets is 539, the number of discrete step of transformation amount, k, varies across the transformations considered. Here, it employed: $k = 14$ for JPEG

compression and uniform light change transformations and $k = 10$ for Gaussian blur. Since the first step of transformation amount corresponds to the reference image, the number of set B_{kd} (Equation 5.4) is 13 for JPEG compression and light changes and 9 for blurring for a total of $2 \times (13 \times 539) + 9 \times 539 = 18865$ repeatability rate for each detector.

5.3.2 Trait Indices

In this section, the trait indices for all the assessed image feature detectors are presented and discussed. The trait indices have been designed to provide a measure of the bias of the feature detector for any of the types of scene introduced by the classification criterion described in the Section 5.2.3. In other words, they are indicative of the types of the scene for which a feature detector is expected to perform well and bad. Accordingly, with the definition provided in the Section 5.2.4, they represent the percentage of the scenes in the top and lowest rankings of a particular type of scene. Thus, they permit to characterise quantitatively the performance of feature detectors from the point of view of the scene content.

The trait indices are built starting from the top and lowest rankings of any feature detector. For obtaining the results presented in this work, the evaluation framework has been applied utilising a ranking length of 20 ($j = 20$). Finally, the related trait indices are computed by applying the equations 5.7 and 5.8 introduced in Section 5.2.4. The results of all detectors are shown in the Figures 5.2–5.12 and discussed in the following sections. The results are presented utilising radar charts: the transformation amounts are shown on the external perimeter and increase clockwise; the trait indices are expressed in percentage values, which increases from the centre (0%) to the external perimeter (100%) of the chart.

5.3.3 EBR trait indices

All the available trait indices of Edge-Based Region (EBR) detector are reported in Figure 5.2. As discussed in Section 3.4, the performances of EBR are very sensitive

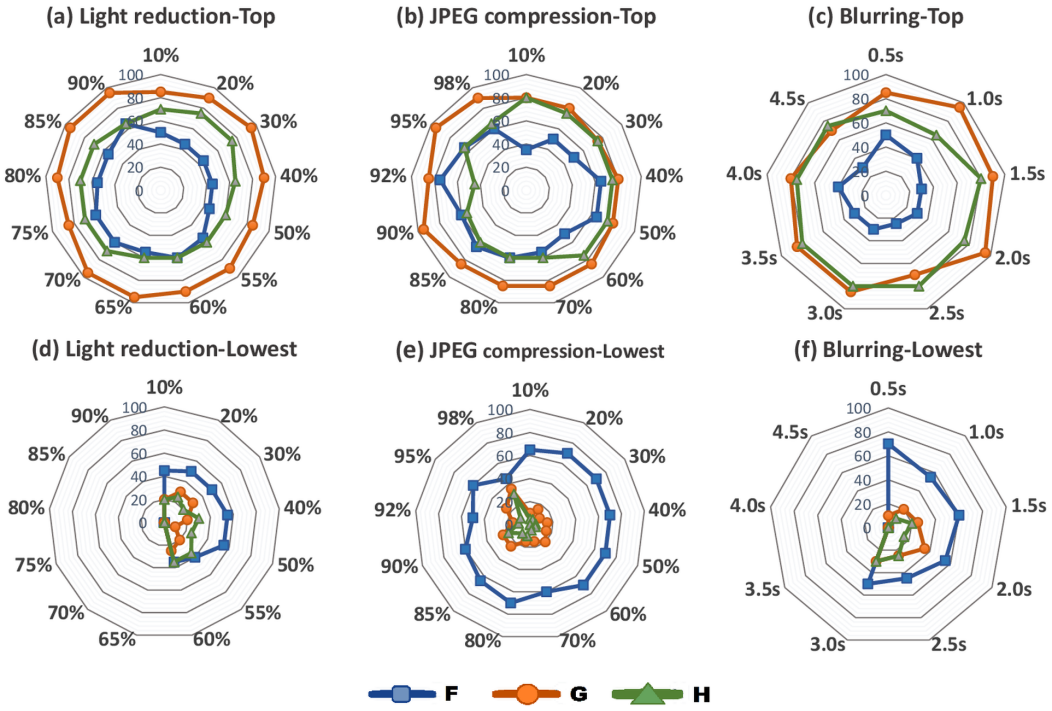

Figure 5.2: Top and lowest trait indices of EBR in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

to uniform light reduction and Gaussian blur so that the number of scenes with a repeatability score equal to 0 is larger than 20 for 65% of light reduction and onward and for blurring higher then 3.0σ. For this reason, the lowest rankings charts are not complete for these two transfomrations (Figure 5.2.d and 5.2.f).

Independently from the transformation type EBR exhibits high values (around 80% - 90%) of G in the top rankings and low values (rarely above 25%) in the lowest rankings denoting a strong bias towards the scenes including many human-made elements. EBR performs generally well on simple scenes as well, in particular under Gaussian blur whose related H index values are never below 70%. The values assumed by F indices are not indicative of the EBR's bias for a particular location type as they assume very similar values between the top and lowest rankings for all the transformations considered.

5.3.3.1 HARLAP and HARAFF trait indices

The rankings of HARLAP and HARAFF are very similar to each other and so are the values of their trait indices. As the EBR detector, both of them are particularly prone

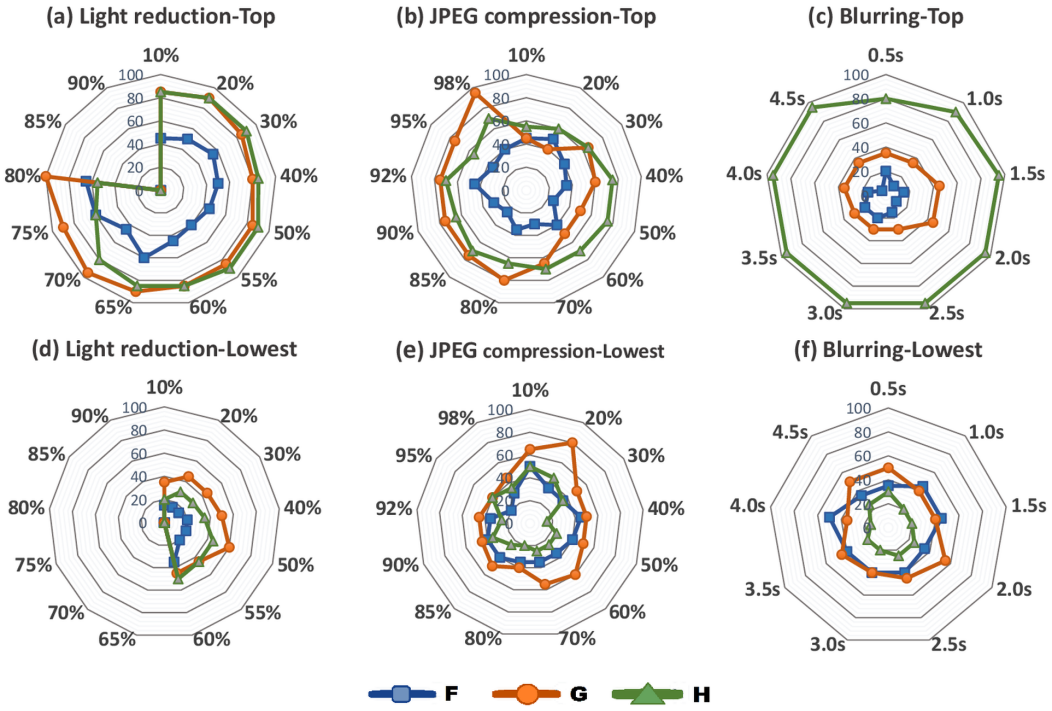

Figure 5.3: Top and lowest trait indices of HARLAP in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

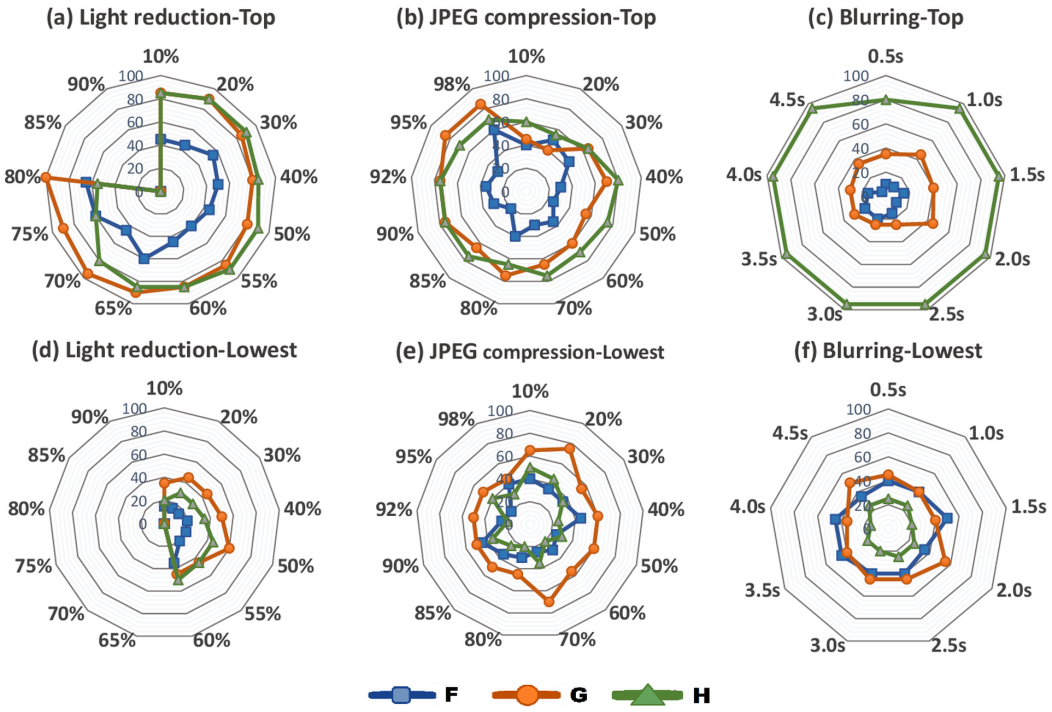

Figure 5.4: Top and lowest trait indices of HARAFF in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

to uniform light changes so the charts are not complete. Figures 5.3 and 5.4 report the results for the top rankings up to 80% of light reduction and up to 60% for the lowest rankings. HARLAP and HARAFF present a bias toward simple scenes, which is particularly strong under uniform light reduction and blurring as can be inferred by the high values that H assumes in the related top twenties. A clear preference of those detectors for human-made objects can be claimed under light changes. However, this is not the case under JPEG compression and Gaussian blur whose related G indices are too close between the top and lowest rankings to draw any conclusion. The F indices are extremely low (never above 20%) for the top twenty rankings under Gaussian blur revealing that HARLAP and HARAFF deal better with non-outdoor scenes under this particular transformation.

5.3.3.2 HESLAP and HESAFF trait indices

Due to the similarities between the method employed for localising the interesting point in images, HESLAP and HESAFF present many similarities between their trait indices. Similarly to HASLAP and HASAFF, uniform light changes have a strong impact on the HESLAP and HESAFF's performance. For that reason, the Figures 5.5 and 5.6 show only the results for the top rankings of up to 80% of light reduction and up to 60% for the lowest rankings.

HESLAP and HESAFF perform better on scenes characterised by simple elements and sharp edges under blurring (especially for high values of σ) and uniform light decreasing. The H indices computed under JPEG compression present fluctuations around 50% for both the top and lowest rankings without bending towards simple nor complex scenes. Both the detectors perform well on scenes containing human-made elements under light reduction, JPEG compression and up to 2.5σ of Gaussian blur. Although both HESLAP and HESAFF do not have any bias for outdoor scenes, the HASLAP's F index appears to be sensitive to blurring variations as it decreases from 45% to 15% constantly as σ increases.

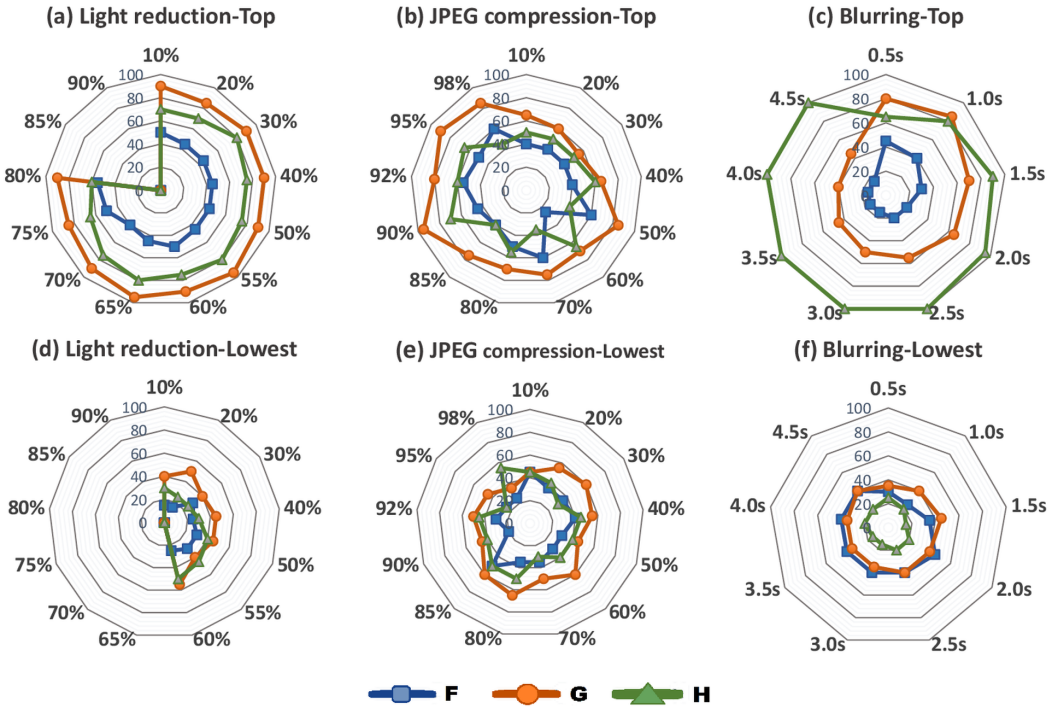

Figure 5.5: Top and lowest trait indices of HESLAP in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

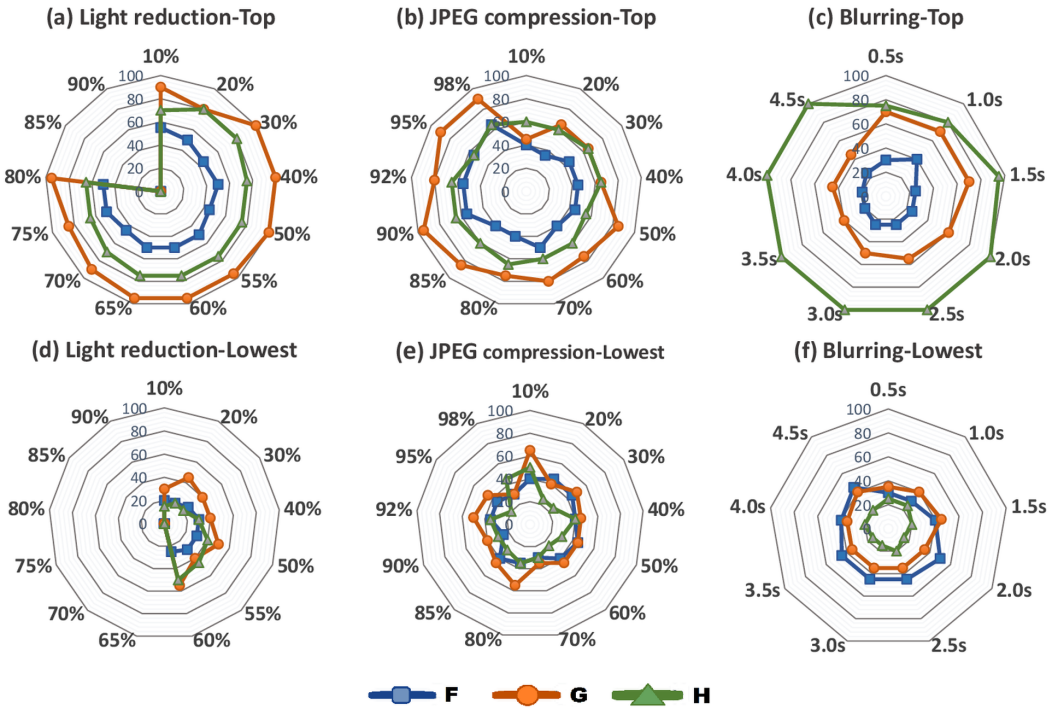

Figure 5.6: Top and lowest trait indices of HESAFF in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

5.3.3.3 SIFT

From the trait indices obtained, it is not possible to determine a clear bias in the performance of SIFT, as their values fluctuate over the entire range of the JPEG compression rate. Figure 5.7 confirms the bias towards simple and human made objects and that bias is stronger at 98% of JPEG compression. However, the indices G and H present large fluctuations in the top twenty scene rankings for the other intermediate compression rates: between 70% and 90% of their values are significantly lower than ones at other compression amounts and reach a minimum at 80% which are 10% for H and 25% for G. Similar variations can be appreciated also for F in both top and lowest rankings with values variations broad up to 40%. While the G and H indices in many cases present small differences between the top and lowest rankings, the F indices are often inversely related. For example, at 30% compression, F is equal to 10% for the top twenty and 60% for the bottom twenty, G is 60% in both cases and H differs for just 20% between the top and lowest rankings.
In conclusion, the classification criteria adopted in this work permits to infer a strong dependency of SIFT from the JPEG compression rate variations, however, it does not allow to draw any conclusions about the general bias, if any exists, towards a particular type of scene.

5.3.3.4 IBR

The uniform light change has a significant impact on the performance of IBR, hence it was not possible to obtain the data for the last two steps of transformation amount for the lowest rankings (Figure 5.8.d). Under light reduction, the presence of a weak bias across all the range of transformation amount is evident for human-made objects: G indices are never below 50% in top rankings while their counterparts in the lowest indices are never above 40%. A similar trend can be observed for F: the share of outdoor scenes in the top twenty is generally below 50%, while is generally never below 50% for light reduction rates from 10% to 65%. Under JPEG compression, IBR achieves better performances on scenes which are both simple and human-made. Indeed, the related G

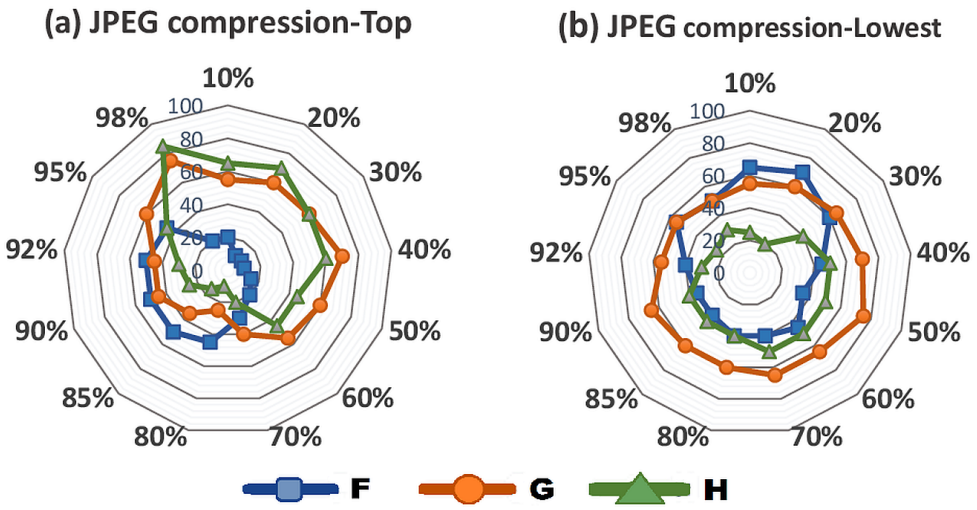

Figure 5.7: Top and lowest trait indices of SIFT in percentage for different amount of JPEG compression (a,b).

and H indices in the top twenties reach very high values, which are never below 75% and 80% respectively (Figure 5.8.b). The same kind of bias observed for JPEG compression characterises IBR under blurring as well: the top rankings are mainly populated by human-made and simple scenes, whereas the lowest rankings contain mostly scenes with the opposite characteristics (Figure 5.8.c and 5.8.f).

5.3.3.5 MSER

The Figure 5.9 shows the trait indices for MSER. Due to sensitivity of MSER to uniform light reduction and Gaussian blur, it has not been possible to compute the trait indices for the lowest rankings at light reductions of more than 60% and for the last three steps of blurring as the number of scenes with repeatability equal to 0 exceed the length of the lowest rankings at those transformation amounts.

The trait indices draw a very clear picture of the MSER's biases. The very high values of G and H of the top ranking indices and the relatively low values obtained for the lowest twenty rankings lead to the conclusion that MSER performs significantly better on simple and human-made dominated scenes for every transformation type and amount. Finally, the outdoor scenes populate mainly the lowest rankings built under light reduction and JPEG compression transformations while F for blurring has low

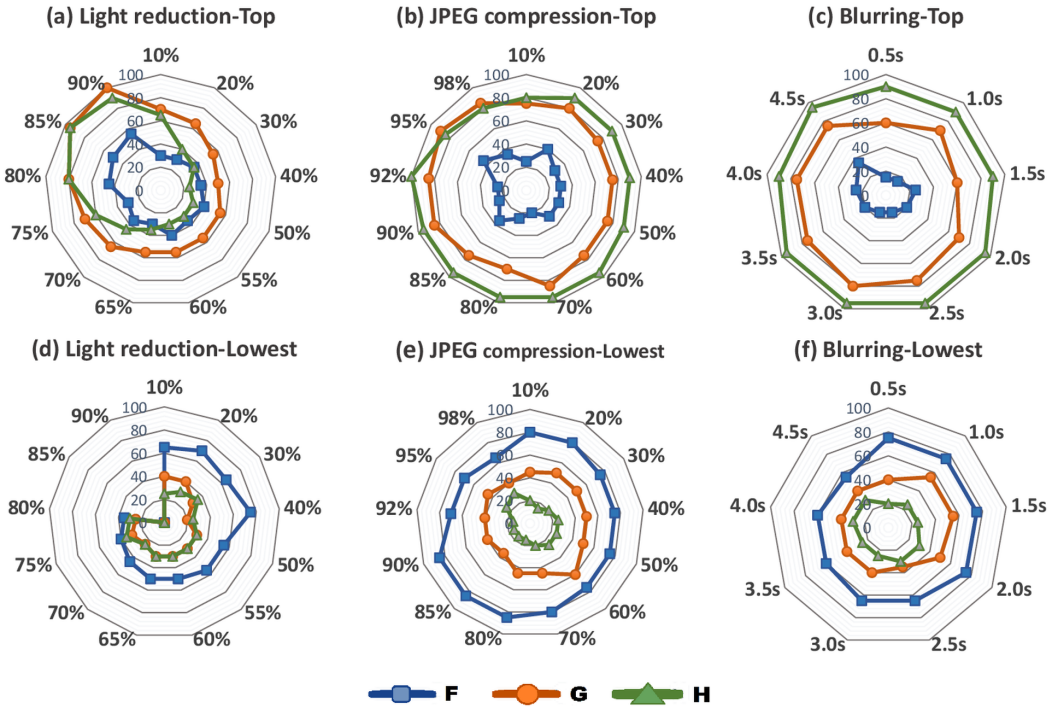

Figure 5.8: Top and lowest trait indices of IBR in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

and balanced values between the top and lowest rankings.

5.3.3.6 SALIENT

The results for uniform light reduction show a strong preference of SALIENT for complex scenes. Indeed, the related H values are low for the top twenty rankings (Figure 5.10.a) and relatively high for the lowest twenty rankings (Figure 5.10.d). Uniform light reduction does not alter the shape of the edges and others lines present in a scene. Thus, the application of the uniform light transformation has the effect of moving scenes whose content remain distinguishable by SALIENT after the reduction of the brightness to the top. These are typically high contrasted scenes characterised by elements whose details remain well discriminable even at the highest amount of light reduction. On the other hand, the results for Gaussian blur (Figures 5.10.c and 5.10.f) show a completely opposite situation in which the most frequent scenes in the top rankings are those characterised by simple structure or, in other words, scenes whose information has relevant components at low frequencies. Indeed, as indicated above, Gaussian blurring can be

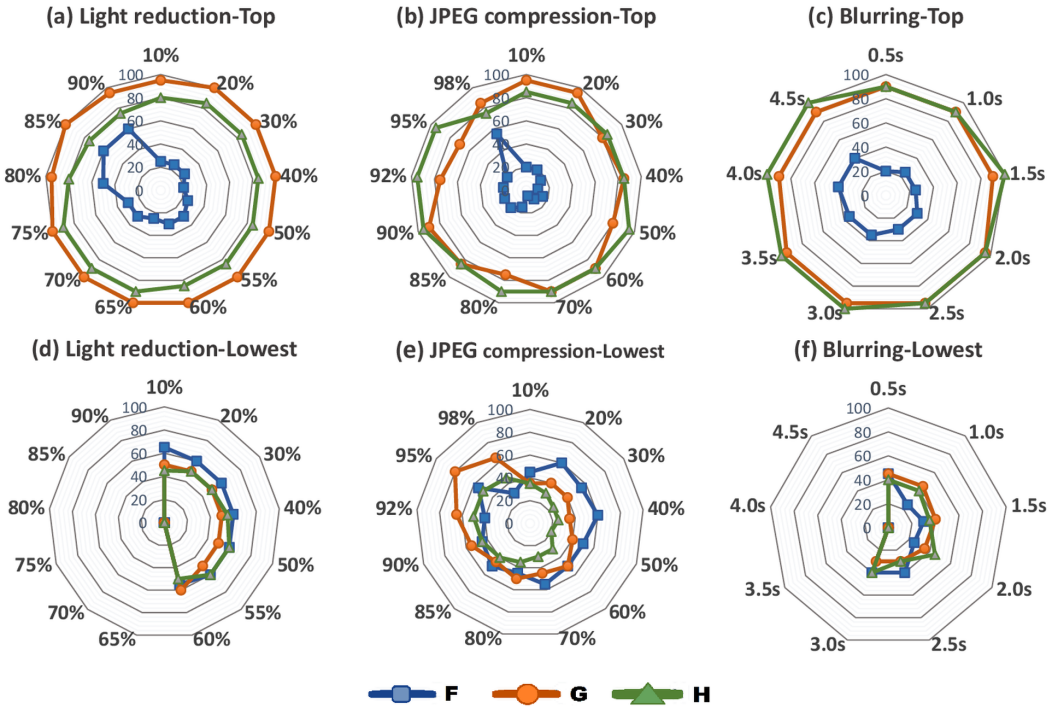

Figure 5.9: Top and lowest trait indices of MSER in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

seen as a low pass filter, so applying it to an image results in a loss of the information at high frequencies. Under JPEG compression, SALIENT exhibits a preference for complex scenes as it is under light reduction with the difference that the H indices increase as the compression rate increases. Although JPEG compression is a lossy process and it may alter the shape of the edges delimiting the potential salient region in an image, the impact on the information content is lighter than the one caused by Gaussian blur. Indeed, the share of simple images is constantly low: H below 30% up to 95%. At 98% the share of simple images in the top twenty increase dramatically to 65% as the images lose a huge part of their information content due to the compression, which produces wide uniform regions as it can be appreciated from the scene sample shown in Figure 3.2.

5.3.3.7 SFOP

Under JPEG compression and Gaussian blur, the bias of SFOP is towards simple scenes representing non-outdoor scenes. The kind of objects favoured is human-made under

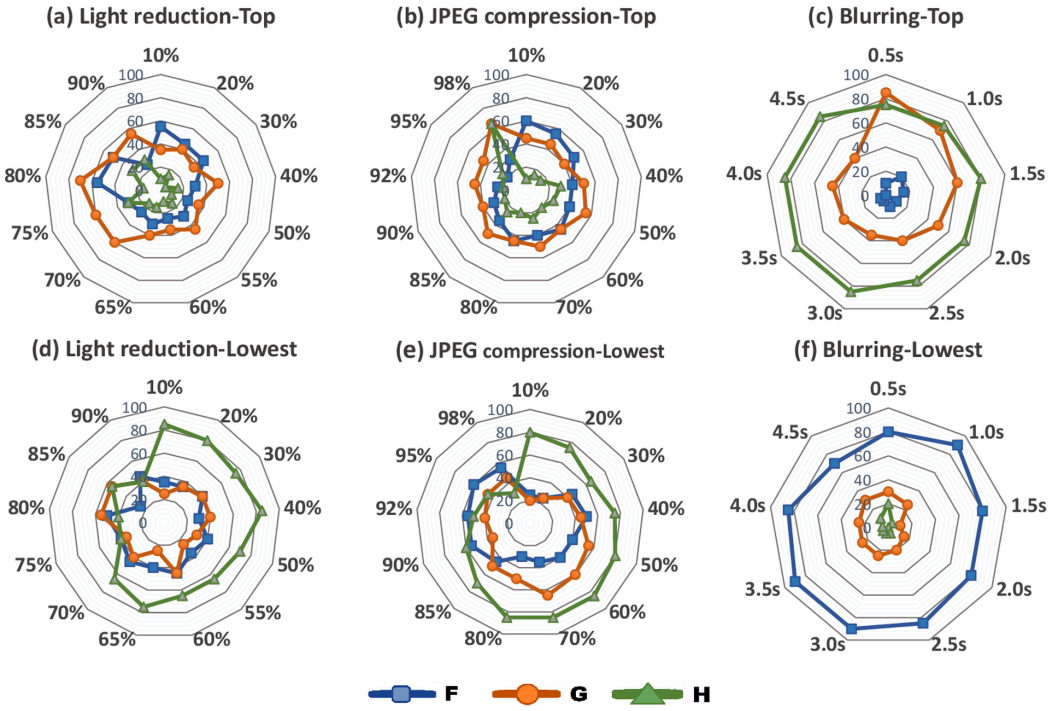

Figure 5.10: Top and lowest trait indices of SALIENT in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

JPEG compression, while for blurring no clear preference can be inferred, due to the closeness of the values of G indices between the top and lowest rankings. The measures of those biases are reflected by the G and H indices shown in Figure 5.11. For both these transformations H assumes high values in the top rankings and low values in the lowest rankings; the indices G for the top rankings of JPEG compression are constantly above 70% whereas the related value registered for the lowest rankings exceed 55% only at 10% of compression rate. The indices obtained for uniform light reduction reveal that SFOP performs worse on outdoor scenes as can be seen from the lowest ranking F values, which mostly fluctuate between 50% and 60% (Figure 5.11.d).

5.3.3.8 SURF

The performance of SURF is particularly affected by uniform light transformation and, because of that, it has not been possible compute the trait indices at 65% and further brightness reductions. The effect of this transformation is to focus the biases of SURF towards human made objects (G greater or equal to 65%). Although the available

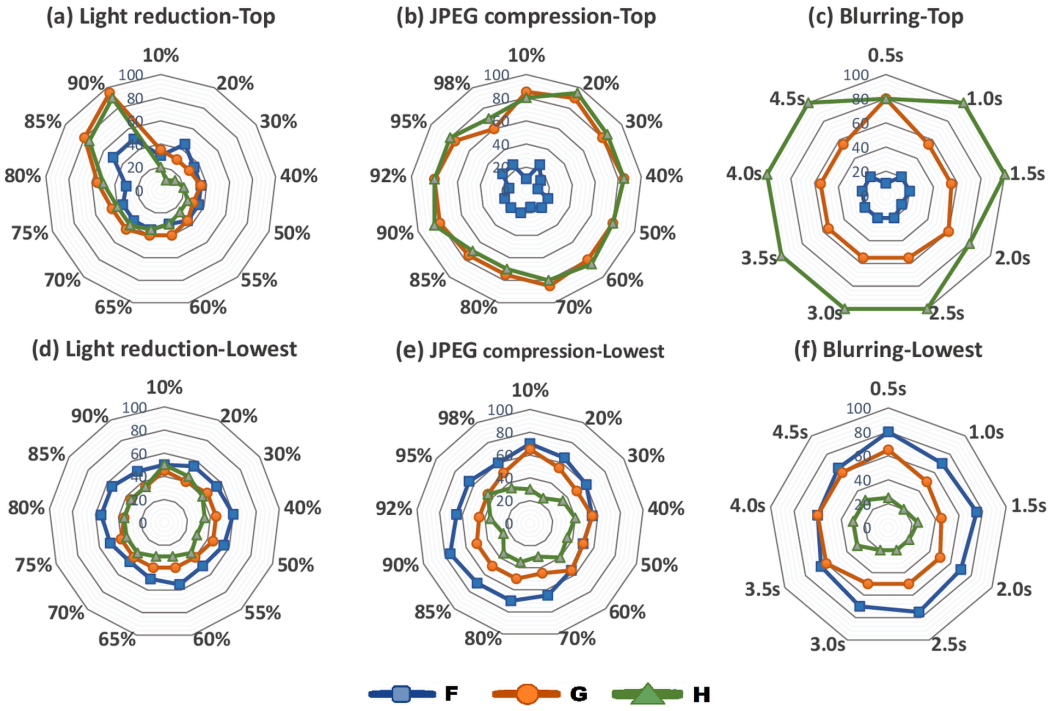

Figure 5.11: Top and lowest trait indices of SFOP in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

F indices for lowest rankings are extremely low (normally within 15%), a weak bias towards outdoor scenes can be claimed as the highest values for F indices in the top twenties reach 60% between 10% and 60% of light reduction. The percentage of simple scenes in the top rankings fluctuates between 50% and 85% which, unfortunately, is reached in a transformation amount range where the indices for lowest rankings are not available, so a comprehensive comparison of their values is not possible. JPEG compression produces more predictable biases on SURF: H's values are significantly higher in the top rankings that in the lowest rankings and the performance are worse with outdoor scenes than with non-outdoor scenes. Finally, the G indices do not express a true bias, neither for human-made nor for natural elements. Under blurring SURF best performs on simple scenes whereas it performs poorer on complex scenes. F values for both rankings groups are very low (except for 0.5σ which reaches 60% for the lowest ranking) while G's values fluctuate around 50%.

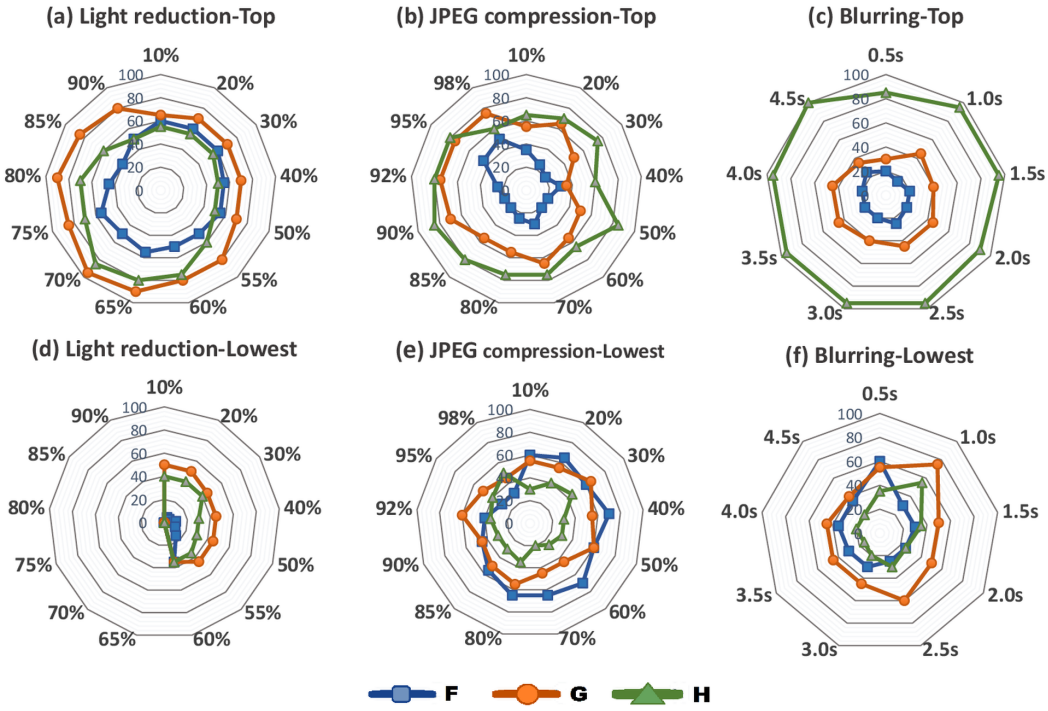

Figure 5.12: Top and lowest trait indices of SURF in percentage for different amount of light reduction (a,d), JPEG compression (b,e) and blurring (c,f).

5.4 Summary

For several state-of-the-art feature detectors, the dependency of the repeatability from input scene type has been investigated utilising a large database composed of images from a wide variety of human-classified scenes under three different types and amounts of image transformations. Although the utilised human-based classification method includes just three independently assigned labels, it is enough to prove that the feature detectors tend to score their highest and lower repeatability scores with particular types of scenes. The detector preferences for a particular category of a scene are pronounced and stable across the type and amount of image transformation for some detectors, such as MSER and EBR. For some detectors, an evident bias emerges only in the presence of a particular transformation and for some of the scene categories such it happens to SALIENT under blurring. In a few cases, the indices do not permit to identify clearly a particular preference of the detector for a particular trait of the scenes. For example, it is unclear to identify if there is a preference for complex or natural objects under

light changes for SFOP. Indeed, its G and H values reveal that both the rankings, top and lowest twenty, include similar shares complex and natural scenes up to 75% of light reduction.

The proposed framework can be utilised for assessing any arbitrary set of detectors. A designer who needs to maximise the performance of a vision system starting from the choice of the better possible local feature detector can take advantage of the proposed framework. Indeed, this framework allows to identify the detectors which perform better with the type of scene most common in an application context before performing any task-oriented evaluation (e.g. [51], [52]). At that point, such a specific evaluation will be carried out on a smaller set of local feature detectors. For example, for an application which has to deal mainly with an indoor environment, the detectors that should be short-listed are HESAFF, HESLAP, HARAFF and HASAFF which have been proven to achieve their highest repeatability rate with non-outdoor scenes. On the other hand, if an application is intended for working in an outdoor environment, EBR should be one of the considered local feature detectors, especially in presence of poor light conditions.

5.5 Limitations

As indicated above, the framework proposed in this chapter allows to analyse the feature detectors' behaviour in relation to scene content and, at the same time, represents a useful tool for facilitating the design of vision applications. However, this framework suffers from a limitation due the human-based scene classification which does not allow its use with automated tools. For example, extending the solution proposed in Chapter 4 in order to take advantage of the trait indices presented in this chapter would require a further stage to determine the scene type. Finding an effective and efficient way to automatically determine the scene content is the direction to follow in order to extend the work presented in this chapter.

6

Conclusions and Future Directions

This chapter presents a summary of the contributions provided by this thesis and suggests some possible directions to extend and improve the research presented in this dissertation.

6.1 Summary of Contributions

This thesis targets the problem of the detection of local features in images, which is a fundamental step of the feature matching process. Although a large number of local feature detectors have been proposed so far, a universal detection method that performs well in any context is not available yet. This research aims to bridge this gap by investigating a solution based on the dynamic selection of the most suitable detector for any given operating condition, which would allow adaptive feature detection stages for vision applications. For this research, predicting how a detector will perform in varying imaging condition is fundamental. The progress made towards a better understanding of the behaviour of local image feature detectors allowed designing a tool for selecting the optimal feature detector in relation to the image transformation, which represents a first but important step towards a fully adaptive selection method of the optimal feature detector for any operation condition.
The contribution provided by this research are summarised below.

- Feature detectors normally operate in complex and variable environments. Consequently, reliable metrics are fundamental in order to predict accurately the performance level a detector will reach operating in real applications. Keeping in mind this, an evaluation framework based on the improved repeatability rate [6] and designed to be employed with large image databases such as [7] has been designed. This framework provides statistically-significant performance indicators to evaluate local feature detectors. Utilising this, several state-of-the-art local feature detectors (such as HARAFF, MSER and SURF) have been assessed under varying JPEG compression, light reduction and Gaussian Blur changes.

- A tool for selecting the optimal feature detector for different amount of JPEG compression, light reduction and Gaussian Blur changes has been designed. In Chapter 4 are presented the results obtained by applying the performance model obtained with the framework from Chapter 3 for several feature detectors. The selection tool presents a good accuracy in choosing the optimal local feature de-

tector. Furthermore, the fast execution time makes this tool suitable for real-time applications.

- The selection tool from Chapter 4 represents a first step towards a fully adaptive feature selection method. However, a more refined selection system would require to include the scene content along with image changes into the performance model of a feature detector. To bridge this gap, in Chapter 5 is introduced a second evaluation framework to investigate detectors' performance in relation to both image transformation and scene content. This framework introduces a binary classification criterion of the scene based on the location type (outdoor or indoor), objects type (natural or human-made) and complexity (complex or simple). A method to represent the tendency of a feature detector to perform well or bad with a particular scene type is also introduced: the trait indices. Using this framework, several local image feature detectors have been assessed under varying JPEG compression, light reduction and Gaussian Blur changes.

6.2 Future Directions

The work presented in this thesis can be extended in different directions. In Chapter 3, a new evaluation framework is presented and several local feature detectors are assessed under varying JPEG compression, Gaussian blur and light reduction. New image databases including a large number of scenes with different transformations such as rotation, scale and viewpoint changes can be generated and utilised with that evaluation framework to obtain how the repeatability is influenced by those image transformations at different amounts.

In Chapter 5 a second evaluation framework that utilise a metric to express the relation between the scene content and repeatability (trait indices) is proposed. This framework assigns labels that can only take two values (e.g. outdoor or not outdoor). This criterion is straightforward to use but does not allow to obtain an accurate classification. For example, a scene containing natural and human-made elements in similar shares cannot be described accurately with those binary labels. In order to better describe such a

scene, non-binary labels that can take more values could be employed instead.

As extensively discussed in Section 5.5, a further and decisive improvement to this second evaluation framework would be introducing a method to determine automatically to which category a scene belongs to. This could be difficult to achieve but it would produce a big improvement for the selection tool presented in Chapter 4 as the scene characteristics could be determined online and included as a part of the selection criterion of the optimal local feature detector.

- THE END -

References

[1] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: a survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3, pp. 177–280, 2008.

[2] K. Mikolajczyk and K. Mikolajczyk, "Scale & affine invariant interest point detectors," *International Journal of Computer Vision*, vol. 60, no. 1, pp. 63–86, oct 2004.

[3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International Journal of Computer Vision*, vol. 65, no. 1-2, pp. 43–72, oct 2005.

[4] S. Ehsan, "Improving the effectiveness of local feature detection," Ph.D. dissertation, University of Essex, 2012.

[5] S. Ehsan, A. F. Clark, B. Ferrarini, N. U. Rehman, and K. D. McDonald-Maier, "Assessing the performance bounds of local feature detectors: Taking inspiration from electronics design practices," in *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on.* IEEE, 2015, pp. 166–169.

[6] S. Ehsan, N. Kanwal, A. Clark, and K. McDonald-Maier, "Improved repeatability measures for evaluating performance of feature detectors," *Electronic Letters*, vol. 46, no. 14, pp. 998–1000, 2010.

[7] S. Ehsan, A. F. Clark, B. Ferrarini, and K. McDonald-Maier, "JPEG, Blur and Uniform Light Changes Image Database." [Online]. Available: http://vase.essex.ac.uk/datasets/index.html

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[10] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, 2005.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.

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

[13] K. Mikolajczyk, "Oxford Data Set." [Online]. Available: http://www.robots.ox.ac.uk/~vgg/research/affine/

[14] K. Rapantzikos, Y. Avrithis, and S. Kollias, "Detecting regions from single scale edges," in *in Proceedings of International Workshop on Sign, Gesture and Activity (SGA'10), European Conference on Computer Vision (ECCV 2010)*, September 2010.

[15] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors," *International Journal of computer vision*, vol. 37, no. 2, pp. 151–172, 2000.

[16] E. Bostanci, N. Kanwal, and A. F. Clark, "Feature coverage for better homography estimation: an application to image stitching," in *Systems, Signals and Image Processing (IWSSIP), 2012 19th International Conference on*. IEEE, 2012, pp. 448–451.

[17] T. Tuytelaars and L. Van Gool, "Content-based image retrieval based on local affinely invariant regions," in *Visual Information and Information Systems*. Springer, 1999, pp. 493–500.

[18] T. Tuytelaars and L. Van Gool, "Matching widely separated views based on affine invariant regions," *International Journal of Computer Vision*, vol. 59, no. 1, pp. 61–85, 2004.

[19] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761–767, 2004.

[20] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," in *ECCV, 2004*, pp. 228–241.

[21] W. Förstner, T. Dickscheid, and F. Schindler, "Detecting interpretable and accurate scale-invariant keypoints," in *IEEE International Conference on Computer Vision*, 2009.

[22] C. Harris and M. Stephens, "A combined corner and edge detector," in *Procedings of the Alvey Vision Conference*. British Machine Vision Association and Society for Pattern Recognition, 1988.

[23] J. Canny, "A computational approach to edge detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, no. 6, pp. 679–698, 1986.

[24] W. Förstner, "A framework for low level feature extraction," in *Proceedings The 3rd European Conference on Computer Vision*, vol. 2. Springer Science & Business Media, 1994, pp. 383–394.

[25] J. Bigün *et al.*, "A structure feature for some image processing applications based on spiral functions," *Computer Vision, Graphics, and Image Processing*, vol. 51, no. 2, pp. 166–194, 1990.

[26] F. Attneave, "Some informational aspects of visual perception." *Psychological review*, vol. 61, no. 3, p. 183, 1954.

[27] D. G. Lowe, "Object recognition from local scale-invariant features," in *IEEE international conference on Computer Vision*, 1999.

[28] A. Haja, B. Jähne, and S. Abraham, "Localization accuracy of region detectors," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on.* IEEE, 2008, pp. 1–8.

[29] T. Dickscheid and W. Förstner, "Evaluating the suitability of feature detectors for automatic image orientation systems," in *Computer Vision Systems.* Springer, 2009, pp. 305–314.

[30] M. Awrangjeb, G. Lu, C. S. Fraser, and M. Ravanbakhsh, "A fast corner detector based on the chord-to-point distance accumulation technique," in *Digital Image Computing: Techniques and Applications, 2009. DICTA'09.* IEEE, 2009, pp. 519–525.

[31] M. Awrangjeb, G. Lu, and C. S. Fraser, "Performance comparisons of contour-based corner detectors," *Image Processing, IEEE Transactions on*, vol. 21, no. 9, pp. 4167–4179, 2012.

[32] N. Sebe, Q. Tian, E. Loupias, M. S. Lew, and T. S. Huang, "Evaluation of salient point techniques," *Image and Vision Computing*, vol. 21, no. 13, pp. 1087–1095, 2003.

[33] M. Awrangjeb, G. Lu, and M. Murshed, "An affine resilient curvature scale-space corner detector," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 1. IEEE, 2007, pp. I–1233.

[34] M. Awrangjeb and G. Lu, "Robust image corner detection based on the chord-to-point distance accumulation technique," *Multimedia, IEEE Transactions on*, vol. 10, no. 6, pp. 1059–1072, 2008.

[35] S. Ehsan, N. Kanwal, A. F. Clark, and K. D. McDonald-Maier, "Measuring the coverage of interest point detectors," in *Image Analysis and Recognition*. Springer, 2011, pp. 253–261.

[36] S. Ehsan, A. Clark, and K. McDonald-Maier, "Rapid online analysis of local feature detectors and their complementarity," *Sensors*, vol. 13, no. 8, pp. 10 876–10 907, aug 2013.

[37] T. Dickscheid, F. Schindler, and W. Förstner, "Coding images with local features," *International Journal of Computer Vision*, vol. 94, no. 2, pp. 154–174, 2011.

[38] H. Aanæs, A. L. Dahl, and K. S. Pedersen, "Interesting interest points," *International Journal of Computer Vision*, vol. 97, no. 1, pp. 18–35, jun 2012.

[39] M. D. Heath, S. Sarkar, T. Sanocki, and K. W. Bowyer, "A robust visual method for assessing the relative performance of edge-detection algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 12, pp. 1338–1359, 1997.

[40] D. Demigny and T. Kamlé, "A discrete expression of canny's criteria for step edge detector performances evaluation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 11, pp. 1199–1211, 1997.

[41] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso, "A comparative evaluation of interest point detectors and local descriptors for visual slam," *Machine Vision and Applications*, vol. 21, no. 6, pp. 905–920, 2010.

[42] M. Asbach, P. Hosten, and M. Unger, "An evaluation of local features for face detection and localization," in *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*. IEEE, 2008, pp. 32–35.

[43] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local features for object class recognition," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 2. IEEE, 2005, pp. 1792–1799.

[44] M. Stark and B. Schiele, "How good are local features for classes of geometric objects," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[45] P. Moreels and P. Perona, "Evaluation of features detectors and descriptors based on 3d objects," *International Journal of Computer Vision*, vol. 73, no. 3, pp. 263–284, sep 2007.

[46] S. Gauglitz, T. Hollerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *International Journal of Computer Vision*, vol. 94, no. 3, pp. 335–360, mar 2011.

[47] F. Crete, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: perception and estimation with a new no-reference perceptual blur metric," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 64 920I–64 920I.

[48] Z. Wang, H. R. Sheikh, and A. C. Bovik, "No-reference perceptual quality assessment of jpeg compressed images," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–477.

[49] P. Tissainayagam and D. Suter, "Assessing the performance of corner detectors for point feature tracking applications," *Image and Vision Computing*, vol. 22, no. 8, pp. 663–679, 2004.

[50] F. Fraundorfer and H. Bischof, "A novel performance evaluation method of local detectors on non-planar scenes," in *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*. IEEE, 2005, pp. 33–33.

[51] M. C. Shin, D. Goldgof, and K. W. Bowyer, "An objective comparison methodology of edge detection algorithms using a structure from motion task," in *Computer*

Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on. IEEE, 1998, pp. 190–195.

[52] M. Shin, D. Goldgof, and K. Bowyer, "Comparison of edge detectors using an object recognition task," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 1. IEEE, 1999, pp. 360–365.