Keypoint descriptor fusion with Dempster-Shafer Theory

V.M. Mondéjar-Guerra, R. Muñoz-Salinas*, M.J. Marín-Jiménez, A. Carmona-Poyato, R. Medina-Carnicer

Maimonides Institute for Biomedical Research (IMIBIC). Department of Computing and Numerical Analysis, Cordoba University, 14071 Cordoba, Spain

Abstract

Keypoint matching is the task of accurately finding the location of a scene point in two images. Many keypoint descriptors have been proposed in the literature aiming at providing robustness against scale, translation and rotation transformations, each having advantages and disadvantages. This paper proposes a novel approach to fuse the information from multiple keypoint descriptors using Dempster-Shafer Theory of evidence [1], which has proven particularly efficient in combining sources of information providing incomplete, imprecise, biased, and conflictive knowledge. The matching results of each descriptor are transformed into an evidence distribution on which a confidence factor is computed making use of its entropy. Then, the evidence distributions are fused using Dempster-Shafer Theory (DST), considering its confidence. As result of the fusion, a new evidence distribution that improves the result of the best descriptor is obtained. Our method has been tested with SIFT, SURF, ORB, BRISK and FREAK descriptors using all possible combinations of them. Results on the Oxford keypoint dataset [2] shows that the proposed approach obtains an improvement of up to 10% compared to the best one (FREAK).

Keywords: Keypoint matching, Local descriptor, Dempster-Shafer

1. Introduction

Keypoint matching is an important task with many applications in areas such as image retrieval [3], image stitching [4], object recognition [5] and stereo matching [6, 7], amongst others. The goal of keypoint matching is to find pixel correspondences representing the same real point in two images. Many local 2D descriptors have been proposed in the last few years for this task, each having its weaknesses and strengths. The most popular ones are SIFT [5], SURF [8], ORB [9], BRISK [10] and FREAK [11]. However, little attention has been paid to the idea of fusing multiple descriptors as to improve their matching capability.

The process of finding and matching the correspondences between two images with 2D descriptors is usually composed by the next three steps. First, keypoints are detected at the most distinctive locations in the images. An ideal detector would detect the same keypoint locations under different image transformations (i.e. rotation, scale, viewpoint or illumination changes). Second, for each detected keypoint a descriptor is built to describe the local region. Finally, keypoints are matched by analyzing the similarity between descriptors in both images. Figure 1 shows a keypoint matching example with two images of the same scene taken from different viewpoints.

This work proposes a novel approach to improve the matching process by fusing the information provided by multiple descriptors using Dempster-Shafer Theory (DST) [1]. DST employs degrees of evidence (a weaker version of probabilities), making the manipulation of uncertainty especially attractive because of its simplicity and because it does not require specifying priors or conditionals that might be unfeasible to obtain in certain problems. In our proposal, each descriptor is employed to generate a sorted list of candidates, which is then transformed into an evidence distribution, from which a global confidence factor is obtained using its entropy. Then, DST is employed to fuse the evidence distributions using the confidence factor to weight the contribution of each descriptor. Finally, the best candidate is selected using the pignistic probability on the fused evidence distribution. Compared to other

☆This work has been developed with the support of the Research Projects called BROCA and TIN2012-32952, both financed by Science and Technology Ministry of Spain and FEDER

*Corresponding author.Tel.:+34-957-21-22-89; fax: +34-957-21-86-30.
Email address: rmsalinas@uco.es (R. Muñoz-Salinas)

Figure 1: Keypoint matching example from the Oxford dataset [2]. Two images of the same scene taken from different angles. Keypoints in both images are represented as circles. Green lines represent correct matches while pink lines are erroneous matches. (Best viewed in color)

approaches, our method does not require a training step determining the performance of each descriptor in order to assign a confidence factor. Instead, the confidence assigned to each descriptor is calculated independently for each keypoint using the entropy of the evidence distribution generated. The experiments conducted on the Oxford dataset [2] show that the proposed method improves the results of the individual components fused, up to a 10% in the best case. In addition, our method compares favorably with the standard fusion approaches presented in [12].

The rest of this paper is organized as follows. After presenting in Section 2 the related works, Section 3 provides a general formulation of the keypoint descriptor fusion problem. Section 4 presents the standard methods for fusion employed in [12] which are compared to our method. Then, Section 5 introduces DST and our proposal. Finally, Sections 6 and 7 present the experiments conducted and draw some conclusions, respectively.

2. Related works

2.1. Keypoint descriptors

In the last years, several local keypoint descriptors have been proposed. The most well-know descriptor is SIFT (Scale-Invariant Feature Transform) by Lowe [5]. It convolves the image with Difference of Gaussians (DoG) functions at multiple scales to detect keypoint locations that are scale-space extremes. The descriptor is built computing several histograms of gradient orientations resulting in a feature vector that is invariant to scale and rotation. It is shown in [13] that SIFT is one of the most effective descriptors but its low speed has been criticized. To deal with this problem Bay *et al.* [8] propose SURF (Speeded Up Robust Features) as a faster descriptor that also maintains a similar matching performance. For the extraction step SURF uses the Fast-Hessian detector, which is based on the determinant of the Hessian matrix. SURF descriptors rely on sums of 2D Haar wavelet responses in (x, y) directions. Many real-time applications demand fast keypoint extractors and low-memory descriptors. As a result, the feature extractor FAST [14](Features from Accelerated Segment Test) and various binary descriptors have been proposed. FAST analyzes a circular region around each keypoint candidate, if there are a minimum number of connected pixels on the region which are brighter or darker than the central pixel the keypoint candidate is selected. Besides the low memory requirements one advantage of binary descriptors is their matching speed: since binary features are employed, the Hamming distance (much faster to compute than the Euclidean one) can be used. Calonder *et al.* [15] introduced the first binary keypoint descriptor called BRIEF (Binary Robust Independent Elementary Features). It computes some binary tests over an image patch and concatenates them into a bit string that compose the descriptor. Each test is a simple binary comparison between the intensity value of two pixel locations belonging to the patch. Later, Rublee *et al.* [9] proposed ORB (Oriented Fast and Rotate BRIEF) as an enhancement of the FAST detector and BRIEF descriptor. It adds a scale pyramid image to create FAST features at different levels and uses the intensity

centroid, computed by standard moments, to obtain the orientation. Then, ORB steers pixels accordingly to build a descriptor that is invariant to rotation. Leutenegger *et al.* [10] proposed BRISK (Binary Robust Invariant Scalable Keypoints) which relies on a circular sampling pattern from which it computes several brightness comparisons creating a binary string invariant to scale and rotation. The last recently remarkable descriptor published is FREAK (Fast Retina Keypoint) by Alahi *et al.* [11] which relies on a retinal sampling pattern, inspired by the human retina, which uses an algorithm to select the set of binary tests that maximizes the diversity between them in order to construct the feature descriptors.

2.2. Keypoint fusion

Despite the many efforts devoted to the development of new keypoint descriptors, little attention has been paid to the idea of fusing the existing keypoint descriptors in order to improve their results. In [16], He *et al.* propose a Multi-Descriptor Multi-Nearest Neighbor, which classifies images by employing its nearest neighbors coming from all of the categories and different kinds of feature descriptors from images. Compared to other NN-based image classification methods, their approach combines different kinds of descriptors with the nearest neighbors coming from all classes instead of using only the K nearest neighbors with only one feature descriptor.

In the work [17], Bakshi *et al.* propose a method for IRIS people recognition that fuses the matches obtained by SURF and SIFT keypoint descriptors. Their approach does the matching with both descriptors independently, and then, analyzes the number of matches obtained by the two descriptors, which is used as a matching score itself.

In the work of Mountney *et al.* [18], the authors propose a Bayesian framework for fusing keypoint descriptors. In their approach, a first training step selects amongst a set of keypoint descriptors those that are most discriminative. Then, a Naive Bayesian Network trained on a subset of data with ground truth is employed to fuse the selected keypoint descriptors. The main drawback of this approach is that it uses a trained scheme which depends on the dataset employed. While a particular keypoint descriptor might not be very informative for a given dataset, it might be useful in an unknown situation, but if it has already been discarded, its contribution cannot be evaluated.

The work of Perakis *et. al* [12] evaluates the fusion of 2D and 3D features in the context of facial landmark detection. In a first step, landmarks are learned from an annotated dataset, and later, a template matching approach is applied for recognition. They evaluate several fusion approaches for template matching showing their advantages and disadvantages. This method is compared with ours in the experimental section.

The work of Weng *et. al.* [19] proposes a feature set matching approach focused on the face recognition problem when there are occlusion or illumination changes. For each keypoint detected, they compute SIFT and SURF descriptors and combine them by simple concatenation. However, as they state, this simple combination cannot be effectively represented in Euclidean space. In order to exploit the discriminative potential of the two concatenated descriptors the authors use Metric Learning [20]. Their experiments show that the combination of SIFT and SURF descriptors can improve the invariance of local features to illumination, viewpoint and pose variations.

In general, the related works require a training step in order to assign weights to the different features. In contrast, we deal with the problem of determining the best match for a source point between a set of target points with no previous knowledge. To that end, our approach creates a set of evidence distributions that are fused using DST.

DST, which is also known as the evidence theory, is a generalization of the Bayes theory of subjective probability. It includes several models of reasoning under uncertainty such as the Smets' Transferable Belief Model (TBM) [21]. It has been applied to several disciplines such as people tracking [22], fraud detection [23], classification [24], risk analysis [25], clustering [26, 27], image processing [28, 29, 30, 31], autonomous robot mapping [32], human-computer interaction [33], land mine detection [34] and driver assistance [35], amongst others.

In [36] Denoeux *et. al.* presents a solution to the multi-target tracking problem. Three attributes (position, velocity, and class) are employed to describe the targets. Then, each attribute is considered a piece of evidence and they are fused using DST. This problem is similar to ours since it is in essence a multi-target assignment problem, where targets are described by multiple attributes.

In the context of classification with DST, entropy has been previously employed as a measure of confidence. In [37], the authors fuse the output of several classifiers for the problem of *word sense disambiguation*. They define the weights of the classifiers adaptively based on ambiguity measures associated with their decisions with respect to each particular pattern under classification. In their approach, the ambiguity is defined by Shannon's entropy. The work of Ranoeliarivao *et. al* [38], tackles the problem of multi-source clustering of remote sensing images based on an unsupervised assignment of evidences. Instead of using a user-defined ambiguity threshold to decide if a sample is

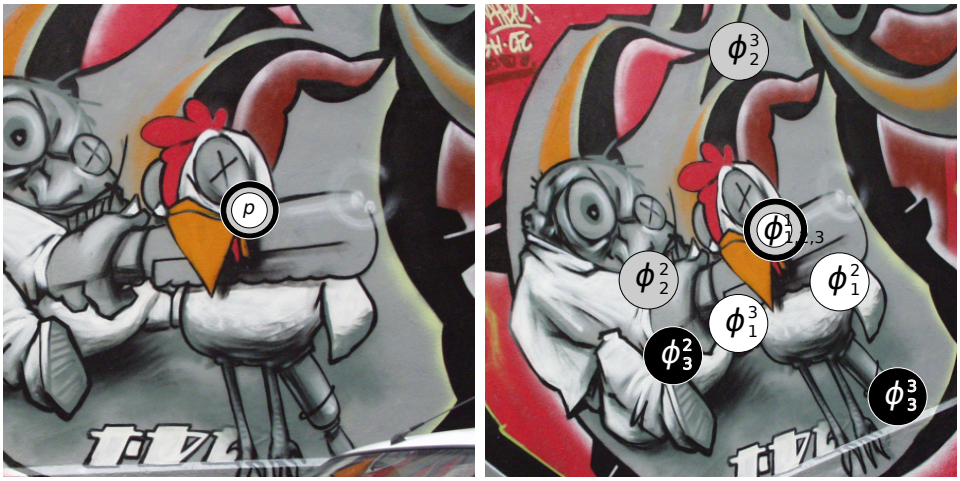

Figure 2: Example of the problem for $K = 3$, where $k = 1$ SIFT, $k = 2$ SURF, $k = 3$ FREAK. (Left) Source image with the keypoint p to be matched. (Right) Target image with the n best matches ϕ_k^i for each descriptor when $n = 3$.

highly ambiguous, they employ an ambiguity factor based on the local entropy among the cluster memberships. In [39], a method to predict ranking preferences is proposed using a discount rate derived from the information entropy. The goodness of the decision-making between different groups of experts is evaluated by analyzing the entropy of their decisions. In our work, the entropy of the evidence distributions is employed as a measure of confidence in the keypoint descriptor.

3. Problem formulation

Let us denote by p a point detected by a keypoint detector in a source image. Our aim is to find the best correspondence q in a target image from the set of keypoints $\mathcal{Q} = \{q^1, \ldots, q^N\}$. It must be noticed that the correspondence might not be in $\mathcal{Q}$ because of occlusion or error of the keypoint detector. Let us denote by

$$\Psi(p) = \{\psi_k(p) \mid k = 1 \ldots K\}, \tag{1}$$

to the set of keypoint descriptors used to describe the point p, where

$$\psi_k(p) = (f_{1p}^k, f_{2p}^k, f_{3p}^k, \ldots f_{n_k p}^k,) \tag{2}$$

is the feature vector provided by the k-th descriptor. Please notice that we are employing a single keypoint detector to extract the keypoints, and then, different descriptors are used on each keypoint.

The first step is to compare the feature vector $\psi_k(p)$ with those of the points in the set $\mathcal{Q}$. Let us define then

$$d_k(p, q) = |\psi_k(p) - \psi_k(q)|, \tag{3}$$

as a distance measure between two keypoint descriptors. Using d_k, the points of $\mathcal{Q}$ are sorted in ascending order independently for each descriptor creating the sets

$$\Phi(p, \mathcal{Q}) = \{\phi_k(p, \mathcal{Q}) \mid k = 1 \ldots K\}, \tag{4}$$

$$\phi_k(p, \mathcal{Q}) = \left\{(\phi_k^1, \ldots, \phi_k^N) \ \in \mathcal{Q} \mid d_k(p, \phi_k^i) \leq d_k(p, \phi_k^j) \ \forall i > j\right\}. \tag{5}$$

The vector $\phi_k(p, \mathcal{Q})$ includes the points from $\mathcal{Q}$ sorted in ascending order according to d_k, i.e., it is a permutation of the elements in $\mathcal{Q}$.

One might expect the first element of the vector $\phi_k(p, \mathcal{Q})$ to be the one with higher likelihood of being the correct match. And ideally, all keypoint descriptors should produce the same ordered vector. However, in practice, some descriptors are better suited for detecting points in particular circumstances than others. Therefore, each vector

$\phi_k(p, \mathcal{Q})$ corresponds to a different permutation of the elements in the set $\mathcal{Q}$. Our goal then is to combine the information in $\Phi(p, \mathcal{Q})$ to find the correct match from the conflicting evidences provided by the different descriptors.

Since the most probable candidates are the first terms of the vector $\phi_k(p, \mathcal{Q})$, using all of them implies an unnecessary increase of memory and computing time. Consequently, we reduce the size of the vectors by retaining only the first n elements, i.e., $|\phi_k(p, \mathcal{Q})| = n$. Figure 2 shows an example with $K = 3$ descriptors keeping the best $n = 3$ matches.

4. Standard feature fusion approaches

This section explains the proposal of Perakis *et. al* [12] for keypoint descriptor fusion, adapted to our work. Although their work is based on facial landmark detection, their approach is a general method valid for any application requiring fusion of descriptors.

In a first step, a descriptor similarity mapping function is employed providing a normalized value indicating how similar two descriptors are. Given the feature vector $\psi_k(p)$, and a target feature vector $\psi_k(q)$, the *normalized distance* $\hat{d}_k(p, q)$ is given by:

$$\hat{d}_k(p, q) = \frac{d_k(p, q)}{d_k(p, \phi_k^n)}. \tag{6}$$

The idea is to divide the distance between p and q by the maximum distance in the set $\phi_k(p, \mathcal{Q})$. Then, a *normalized similarity measure* can be defined by using the function:

$$S_k(p, q) = 1 - \hat{d}_k(p, q)^e, \tag{7}$$

where e favors the nearest elements to the target. When $e = 1$ we obtain a linear mapping, while using $e = 2$ we obtain the quadratic mapping (as expressed in [12]). The similarity measure $S_k(p, q)$ is one if both feature vectors are equal, and is 0 for the farthest element.

The similarity measures obtained for the K keypoint descriptors can be fused by using any of the following rules:

- Sum rule:

$$S_A(p, q) = \frac{1}{N} \sum_{k=1}^{K} S_k(p, q). \tag{8}$$

- Root-mean-square rule:

$$S_E(p, q) = \frac{1}{\sqrt{K}} \left(\sum_{k=1}^{K} S_k(p, q)^2 \right)^{1/2}. \tag{9}$$

- Product rule:

$$S_G(p, q) = \left(\prod_{k=1}^{K} S_k(p, q) \right)^{1/K}. \tag{10}$$

- Max rule:

$$S_{max}(p, q) = max_{k=1}^{K} \left(S_k(p, q) \right). \tag{11}$$

Each of these rules causes a different fusion behavior. In this work, we are testing different combinations of normalized similarity measures and fusion rules so as to find the best one for our problem.

5. Dempster-Shafer proposal for keypoint fusion

This section explains our proposal for keypoint descriptor fusion using DST. First, we provide a brief introduction to DST, and then, we detail our fusion approach.

5.1. Dempster-Shafer theory of evidence

Let us consider a variable ω taking values in the *frame of discernment* Ω, and let us denote to *the power set* 2^Ω. A basic belief assignment (bba)

$$m : 2^\Omega \rightarrow [0, 1],$$

is a function that assigns masses of belief to the subsets A of the power set, verifying:

$$\sum_{A \in \Omega} m(A) = 1. \tag{12}$$

While the evidence assigned to an event in the Bayesian approach must be a probability distribution function, the mass $m(A)$ of a power set element can be a subjective function expressing how much evidence supports the fact A. Furthermore, complete ignorance about the problem can be represented by $m(\Omega) = 1$.

The original Shafer's model imposes the condition $m(\emptyset) = 0$ in addition to that expressed in Eq. 12, i.e., the empty subset should not have mass of belief. However, Smets' TBM model relaxes that condition so that $m(\emptyset) > 0$ stands for the possibility of incompleteness and conflict [21]. In the first case, $m(\emptyset)$ is interpreted as the belief that something out of Ω happens, i.e. accepting the *open-world assumption*. In the second case, the mass of the empty set can be seen as a measure of conflict arising when merging information from sources pointing towards different directions. Nonetheless, a renormalization can transform a Smets' bba m into a Dempster's bba m^* as:

$$\begin{array}{l} m^*(\emptyset) = 0, \\ m^*(A) = \frac{m(A)}{1-m(\emptyset)} \quad if\ A \neq \emptyset. \end{array} \tag{13}$$

A mass function m may be represented in several equivalent ways, of which *commonality* is of importance in this work:

$$q(A) = \sum_{B \supseteq A} m(B),\ \forall A, B \subseteq \Omega. \tag{14}$$

Using the above concepts, it is possible to define the generalized simple bba [40] as a function μ from 2^Ω to $\mathbb{R}$ verifying

$$\begin{array}{l} \mu(A) = 1 - w(A) \\ \mu(\Omega) = w(A) \\ \mu(B) = 0\ \forall B \in 2^\Omega \setminus \{A, \Omega\}. \end{array} \tag{15}$$

for some $A \neq \Omega$ and some $w \in [0, +\infty)$. Let us denote such a mass function as A^w. The weights $w(A)$ can be obtained from the commonalities using the following formula:

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}. \tag{16}$$

This is another alternative representation of a bba which is very useful for fusing different pieces of evidences. Obtaining the masses back from the weights is possible using the Fast Möbius Transform [41].

In this work, we employ the Frank's family of t-norms [42] :

$$x \top_s y = log_s \left(1 + \frac{(s^x - 1)(s^y - 1)}{s - 1}\right), \tag{17}$$

where log_s represents the logarithm function with base $s > 0$. Each value of s defines a t-norm. As $s \rightarrow 0$, Eq. 17 tends to the minimum between x and y, and to the product as $s = 1$. Using this terminology, the combination of two bbas is defined as:

$$m_1 \textcircled{\scriptsize T}_s\ m_2 = \bigcirc\!\!\!\!\!\!\bigcap_{A \in \Omega} A^{w1(A) \top_s w_2(A)}, \tag{18}$$

which is a flexible way to modify with the parameter s the properties of the combination rule. When $s = 0$, the Denoeux [43] cautious rule is obtained, which implements the Least Commitment Principle (LCP). It postulates that, given two belief functions compatible with a set of constraints, the most appropriate is the least informative. This is a rule which assumes that the sources manage dependent pieces of information. On the other hand, when $s = 1$, we obtain Dempster's conjunctive sum operation [44] which can be equivalently defined as:

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \quad \forall A \subseteq \Omega. \tag{19}$$

This rule assumes sources are uncorrelated.

In some applications it is necessary to make a decision and choose the most reliable single hypotheses ω. To do so, Smets [45] proposed the use of the pignistic transformation that is defined for a normal bba as:

$$BetP(\omega) = \sum_{A \subseteq \Omega, \omega \in A} \frac{m(A)}{|A|}. \tag{20}$$

Other approaches for that end are the intersection probability [46], the orthogonal projection and the relative plausibility [47].

5.2. Proposed approach: Keypoint descriptor fusion with DST

Using as input the vectors of $\Phi(p, \mathcal{Q})$ (i.e., each $\phi_k(p, \mathcal{Q})$), we construct the evidence distribution set

$$\Pi(p) = \{\pi_k(p) \mid k = 1 \dots K\}, \tag{21}$$

such that

$$\pi_k(p) = \left(\delta_k(p, \phi_k^1), \dots, \delta_k(p, \phi_k^n)\right), \tag{22}$$

where

$$\delta_k(p, q) = \begin{cases} \varphi \left(\frac{d_k(p, \phi_k^1)}{d_k(p, q)} \right)^{\beta} & if\ q \in \phi_k(p, \mathcal{Q}) \\ 0 & otherwise. \end{cases} \tag{23}$$

The value ϕ_k^1 is the first element of $\phi_k(p, \mathcal{Q})$, φ is a constant factor so that the integral of the distribution is one, and β is an exponential parameter (experimentally evaluated in Sect. 6) that modifies the shape of the distribution.

The evidence distributions created with Eq.22 have the following properties. First, the best candidate (the one with smallest distance) receives the highest evidence. Second, it is a normalized distribution that can be used to compare the results of different keypoint descriptors. It must be reminded that each descriptor works in its own feature space (with its own number of features), thus, distances between different descriptors are not directly comparable. Finally, the parameter β is employed to modify the distribution. Values of β greater than one tend to increase the influence of the first terms, while values smaller than one tend to smooth the distribution.

An ideal descriptor would produce a very small distance for the correct candidate and a very high distance for the rest. So, an ideal descriptor would produce an evidence distribution with a very high value for the first element and very low values for the rest. However, if the descriptor is not good, it might provide high (or low) distance values for all keypoints, including for the correct match, thus producing an uniform distribution. Finally, if the source keypoint is part of a repetitive pattern (such as a chessboard), even a good descriptor would have many possible correct matches, again producing an uniform distribution.

Based on these ideas, the Shannon's entropy of the evidence distribution is proposed as confidence factor, which is calculated as:

$$c_k(p) = 1 - \sum_{i=1}^{n} -\delta_k(p, \phi_k^i) \log\left(\delta_k(p, \phi_k^i)\right). \tag{24}$$

When the entropy of the distribution is 0 (i.e., there is only one element with evidence 1 and the rest is zero) then, $c_k(p)$ is one. As the evidence distribution $\pi_k(p)$ tends to an uniform, $c_k(p)$ tends to zero.

The evidence distributions $\pi_k(p)$ and the confidence values $c_k(p)$ are employed to define the bba for each keypoint descriptor. In our problem, the set of facts $\mathcal{S}$ is comprised by all the points from $\mathcal{Q}$ selected to create the evidence distributions. Thus, we can define $\mathcal{S}$ as the set of points from $\mathcal{Q}$ that are in $\Phi(p, \mathcal{Q})$:

$$\mathcal{S} = \{\mathrm{w} \in \mathcal{Q} \mid \exists k,\ \mathrm{w} \in \phi_k(p, \mathcal{Q})\}. \tag{25}$$

Please notice that $\mathcal{S}$ is a subset of $\mathcal{Q}$. Considering each keypoint descriptor as an independent sensor, the bba of the k-th sensor is defined from Eqs 23 and 24 as:

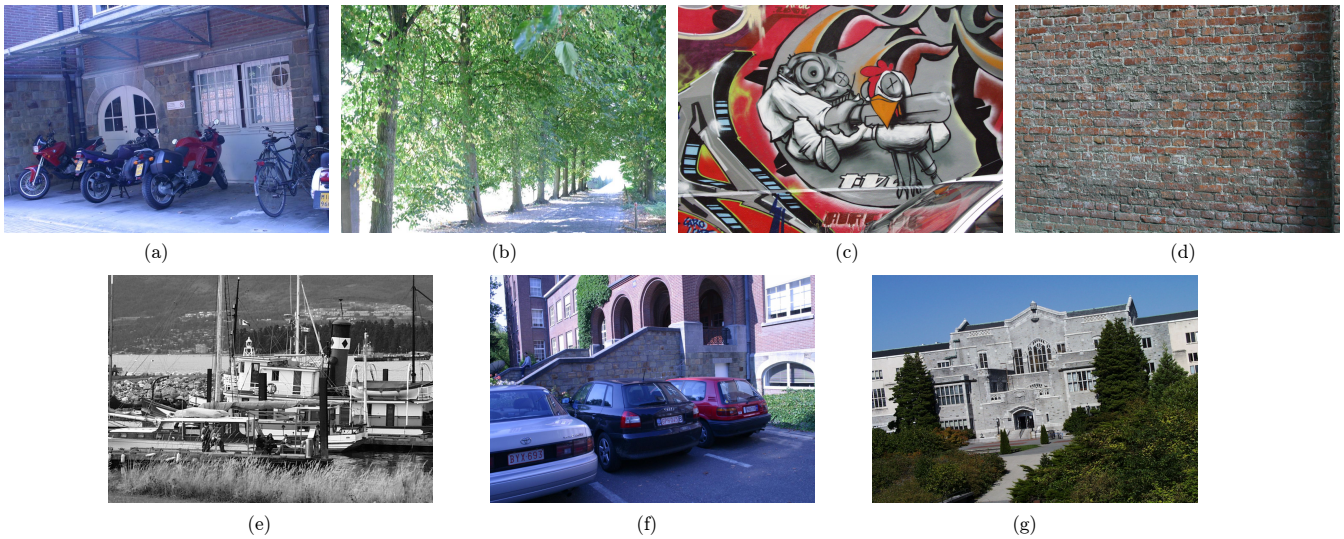

Figure 3: Images from Oxford dataset: blur (a)(b), viewpoint change (c)(d), zoom and rotation (e), lighting change (f) and JPEG compression (g)

$$\begin{aligned} m_k(\Omega) &= c_k(p), \\ m_k(\mathrm{w}) &= \delta_k(p, \mathrm{w}). \end{aligned} \tag{26}$$

Using the t-norm combination rules previously explained (Eq. 17), the fused masses of belief over the power set $\mathcal{S}$ are obtained. Finally, the pignistic transformation (Eq.20) is applied on the fused results sorting the keypoints of $\mathcal{S}$ in descending order:

$$\mathcal{S}' = \left(\mathrm{w}^1, \mathrm{w}^2, \ldots, \mathrm{w}^{|\mathcal{S}|} \mid \mathrm{w}^i \in \mathcal{S} \ \wedge \ BetP(\mathrm{w}^i) > BetP(\mathrm{w}^j) \ \forall i < j\right). \tag{27}$$

In other words, the keypoint w^1 is the most likely keypoint in $\mathcal{Q}$ to be the correct match, w^2 the second most likely and so on. It must be noted that conflict between the descriptors has not been managed at this level. Instead, the ratio between the two best descriptors ($\mathbf{w}^2/\mathbf{w}^1$) is later analyzed to see if the match can be reliably considered, by ensuring it is below a threshold α, thus preventing false positives. This is better explained in the next Section.

6. Experiments and results

This section presents the experiments conducted to test the proposed method. The well-know Oxford dataset[1] of Schmid and Mikolajczyk [2] has been employed in this work for evaluation purposes. It consists of seven scenes that include different image transformations: zoom and rotation (Boat), image blur (Bikes and Trees), view-point change (Graffiti and Wall), light change (leuven) and JPEG compression (ubc), see Fig. 3. Each scene contains a sequence of five images sorted by the amount of transformation. The homography H with respect to the first image that is used to determine the ground truth is also provided by the dataset. The matching comparatives are performed against the first sequence image in all cases, i.e., the first image plays the role of source image, and the rest are the target images.

For all tested methods, the nearest neighbor distance ratio matching (NNDR) [13] strategy has been applied. It means that a possible match provided by a descriptor is only selected if it satisfies the condition $d(p, \phi_k^1)/d(p, \phi_k^2) < \alpha$, where $\alpha \in [0, 1]$. This approach prevents false positives when a descriptor provides similar responses to the two best matching candidates. In our method, the same result is obtained as the ratio between the two most likely candidates, i.e., $(\mathbf{w}^2/\mathbf{w}^1) < \alpha$.

In order to evaluate if a match is correct, the ratio of the intersection over union [13] is applied. In short, a match is considered correct if the overlap error in the image area covered by two corresponding regions ϵ is less than the 50%

[1]The dataset is available at `http://www.robots.ox.ac.uk/~vgg/research/affine/`

Kp-Detector	Kp-Descriptor	F measure
SURF	**SIFT-L1**	**0.721**
SURF	SIFT-L2	0.726
SURF	**SURF-L1**	**0.629**
SURF	SURF-L2	0.559
SURF	**ORB**	**0.570**
SURF	**BRISK**	**0.688**
SURF	**FREAK**	**0.755**

Table 1: Baseline performance of individual keypoint descriptors (Kp) in the database when using the SURF keypoint detector. Lines in bold font are selected for comparison.

Kp-Detector	Kp-Descriptor	F measure
SIFT	SIFT-L1	0.451
SIFT	SIFT-L2	0.437
SIFT	SURF-L1	0.188
SIFT	SURF-L2	0.167
SIFT	ORB	0.369
SIFT	BRISK	0.377
SIFT	FREAK	0.354

Table 2: Baseline performance of individual keypoint descriptors in the database when using the SIFT keypoint detector.

of the region union. For that purpose, the parameter ϵ is calculated as:

$$\epsilon = 1 - \frac{P \cap H^T \cdot Q \cdot H}{P \cup H^T \cdot Q \cdot H}. \tag{28}$$

where P and Q are the regions centered on the source point p and the target point matched q, and H is the homography that relates source and target images.

6.1. Evaluation methodology

In order to test our proposal, the following methodology has been applied. First, keypoints have been selected using a keypoint detector (both SURF and SIFT have been tested). Afterwards, the five keypoint descriptors SIFT, SURF, ORB, BRISK and FREAK have been applied to the detected keypoints. Using the same keypoint detector for all descriptors ensures that the results are comparable. The list of keypoints detected, their descriptors as well as the source code employed in this work are available as supplementary material of this paper for evaluation purposes[2].

The precision, recall and F measure of each descriptor have been evaluated in the dataset, providing the individual baseline performance. These three measures are defined over the true positives (tp), the false positives (fp), and the false negatives (fn) matches as:

$$precision = \frac{tp}{tp + fp}; recall = \frac{tp}{tp + fn}; F = 2\frac{precision \cdot recall}{precision + recall}. \tag{29}$$

Precision is a measure indicating how good a method is at not selecting incorrect matches, while recall measures how many good matches a method is capable of finding. These two measures complement to each other. Finally, the F measure is the harmonic mean of both providing an unique score about the performance of a method. We will focus our analysis on the F measure since it is a unique indicator to determine the performance of a method.

In order to evaluate the performance of the different fusion methods, the non parametric Wilcoxon signed rank test [48] has been applied on the F measure. The test evaluates two methods by considering two hypotheses: the null hypothesis H_0, which assumes that both methods presents similar results, and the alternative hypothesis H_1, indicating that there exists statistically significant differences between the observed results. In all our tests, the value $p = 0.05$ has been employed, indicating that results of the test are reliable at 95.5%.

6.2. Baseline performance

At this point, each descriptor is evaluated individually so as to later analyze the improvements produced by fusion. The values presented in Tables 1 and 2, are the results of each descriptor obtained for the best value of α over the entire dataset using values in the range $[0, 1]$ at regular intervals of 0.01. In other words, each descriptor has been applied to all the sequences for varying values of α, and the solution with the highest F measure is the one shown in the Table. The first and second columns indicate the keypoint detector and descriptor employed, respectively. For SURF and SIFT descriptors, both the L1 and L2 norms have been employed. Since the other descriptors are binary, the Hamming distance has been employed.

[2]The supplementary material is available at `http://www.uco.es/investiga/grupos/ava/node/48/`

First, it can be seen that the SURF keypoint detector provides better performance in all cases. Thus, this will be the one selected for the rest of the experiments. It can also be observed that, when using the SURF keypoint detector, the FREAK descriptor is the best one, while ORB is the worst one. Regarding the distance employed, it can be seen that $L2$ causes a performance reduction in most cases. The options presented in bold font in Table 1 are the descriptors employed for the rest of our comparisons.

6.3. Standard feature fusion performance

The following section evaluates the performance of the standard feature fusion methods presented in Sect. 4. In total, we have tested two normalized similarity measures ($e = \{1, 2\}$ of Eq. 7) and four fusing rules (sum, root-mean-square, product and max in Eq. 8-11). In total, the eight different fusion approaches have been tested, and for each one, all possible combinations of the five descriptors have been tested. The results are presented in Table 3, where the asterisks indicate the descriptors fused. Again, the results reported are those obtained for the best value of α over the entire dataset using values in the range $[0, 1]$ at regular intervals of 0.01.

For instance, line 3, column *($e = 1$,max)* presents the result of fusing all the descriptors but ORB using a linear mapping ($e = 1$) and the max fusion rule which obtain $(+0.30\mathbf{E}^{-2}, \mathbf{H}_0)$. In this case, the fusion method results in an increase of the F measure of $+0.30\mathbf{E}^{-2}$ compared to the best individual descriptor of the four employed, which in this case is FREAK. However, the value $\mathbf{H}_0$ means that the Wilcoxon test [48] indicates that the result is not statistically significant. In other words, the two methods can be considered equivalent in performance. The last line (26), summarizes the results of all the tests performed for a method. So, for the case *($e = 1$,max)*, the result $(-1.92\mathbf{E}^{-2}, \mathbf{H}_1)$ indicates that, considering all possible combinations, the fusion method, in average, obtains a performance reduction and that it is a statistically significant result. So, it can be said that, overall, this fusion method is worst than using the best individual descriptor.

If we analyze the results for the rest of the fusion methods, it can be seen that no one obtains better results that the use of individual classifiers.

6.4. Evidential feature fusion performance

This section evaluates the performance of the proposed approach. As in the previous case, our method has been evaluated by selecting all possible combinations of the five descriptors. Our method has three parameters: n (the cardinal of $\phi_k(p, \mathcal{Q})$ in Eq. 5), β from Eq. 23 and the value of s for the t-norm employed (Eq. 17). The first parameter is the number of keypoints that are employed to create the evidence distribution. The values $n = (4, 3, 2)$ have been employed in our tests. The second parameter β is the factor that modulates the shape of the evidence distribution. Values in the range $\beta = [1, 15]$ at regular intervals of 0.25 have been tested. For the t-norm, the values $s = \{0, 0.5, 1\}$ have been employed. Finally, we have employed values of α in the range $[0, 1]$ at regular intervals of 0.1. The results obtained are presented in Table 4.

As can be observed, the proposed approach obtains better results than the methods previously tested. When using the t-norms $s = \{0, 1\}$ (Cautious and Conjunctive rules), our method always improves the results of the best individual classifier. In general, the conjunctive rule obtains slightly better results than the cautious one, presumably because information managed by the descriptors cannot be considered highly correlated. We would like to draw attention on line 3 and column $s = 1$, where our methods fuses all descriptors but ORB. In that case, we increase performance ($+7.17\mathbf{E}^{-2}$) with respect to FREAK, that scored 0.775. So, in relative terms, it is an improvement of nearly 10% using our approach. In general, the best results are obtained when the ORB keypoint descriptor is not employed.

6.4.1. Influence of the parameters in the method performance

The results previously reported have shown the performance of our method for the best configuration of the parameters n and β. Below, we analyze how the results of our method vary as these parameters change. First, the influence of n has been evaluated in each case. For that purpose, we have run a Wilcoxon test comparing the results of the best configuration of β for the values of $n = (2, 3, 4)$. The results are in Table. 5. While each row of the table shows one fusion combination, the last three columns show the result of the Wilcoxon test comparing the best configuration in terms of F measure. For instance, column "2 vs 3" of line 1 compares the best matcher found using $n = 2$ with the best matcher using $n = 3$ when our method is applied fusing all descriptors but FREAK. The Wilcoxon test indicates that while the approach using $n = 2$ vs $n = 3$ obtains an increment of 0.12% in terms of F measure, the differences observed are not statistically significant. Analyzing the rest of the tests, it can be observed that there is no clear

	SIFT	SURF	ORB	BRISK	FREAK	($e=1$,max)	($e=1$, product)	($e=1$,rms)	($e=1$,sum)	($e=2$,max)	($e=2$,product)	($e=2$,rms)	($e=2$,sum)
0	*	*	*	*	*	$(-0.76\mathbf{E}^{-2},\mathbf{H}_0)$	$(-15.71\mathbf{E}^{-2},\mathbf{H}_1)$	$(+0.42\mathbf{E}^{-2},\mathbf{H}_0)$	$(+1.41\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.21\mathbf{E}^{-2},\mathbf{H}_0)$	$(-13.68\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.04\mathbf{E}^{-2},\mathbf{H}_0)$	$(+1.07\mathbf{E}^{-2},\mathbf{H}_0)$
1	*	*	*	*		$(+0.20\mathbf{E}^{-2},\mathbf{H}_0)$	$(-8.50\mathbf{E}^{-2},\mathbf{H}_1)$	$(+1.45\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.72\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.83\mathbf{E}^{-2},\mathbf{H}_0)$	$(-7.93\mathbf{E}^{-2},\mathbf{H}_1)$	$(+0.16\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.19\mathbf{E}^{-2},\mathbf{H}_0)$
2	*	*	*		*	$(-1.63\mathbf{E}^{-2},\mathbf{H}_0)$	$(-14.80\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.34\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.28\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.04\mathbf{E}^{-2},\mathbf{H}_0)$	$(-13.24\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.10\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.17\mathbf{E}^{-2},\mathbf{H}_0)$
3	*	*		*	*	$(+0.30\mathbf{E}^{-2},\mathbf{H}_0)$	$(-9.18\mathbf{E}^{-2},\mathbf{H}_1)$	$(+0.94\mathbf{E}^{-2},\mathbf{H}_0)$	$(+1.86\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.70\mathbf{E}^{-2},\mathbf{H}_0)$	$(-8.23\mathbf{E}^{-2},\mathbf{H}_1)$	$(+0.45\mathbf{E}^{-2},\mathbf{H}_0)$	$(+1.60\mathbf{E}^{-2},\mathbf{H}_1)$
4	*		*	*	*	$(-1.46\mathbf{E}^{-2},\mathbf{H}_0)$	$(-12.08\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.00\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.60\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.15\mathbf{E}^{-2},\mathbf{H}_0)$	$(-10.52\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.48\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.45\mathbf{E}^{-2},\mathbf{H}_0)$
5		*	*	*	*	$(-1.84\mathbf{E}^{-2},\mathbf{H}_0)$	$(-11.90\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.01\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.94\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.05\mathbf{E}^{-2},\mathbf{H}_1)$	$(-11.19\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.15\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.61\mathbf{E}^{-2},\mathbf{H}_0)$
6	*	*	*			$(-1.60\mathbf{E}^{-2},\mathbf{H}_0)$	$(-7.71\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.89\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.24\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.45\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.74\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.64\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.33\mathbf{E}^{-2},\mathbf{H}_0)$
7	*	*		*		$(+1.01\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.14\mathbf{E}^{-2},\mathbf{H}_1)$	$(+1.12\mathbf{E}^{-2},\mathbf{H}_0)$	$(+1.28\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.15\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.02\mathbf{E}^{-2},\mathbf{H}_1)$	$(+0.50\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.58\mathbf{E}^{-2},\mathbf{H}_0)$
8	*	*			*	$(-1.12\mathbf{E}^{-2},\mathbf{H}_0)$	$(-7.94\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.90\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.14\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.48\mathbf{E}^{-2},\mathbf{H}_0)$	$(-7.05\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.28\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.04\mathbf{E}^{-2},\mathbf{H}_0)$
9	*		*	*		$(-1.00\mathbf{E}^{-2},\mathbf{H}_0)$	$(-5.82\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.43\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.41\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.11\mathbf{E}^{-2},\mathbf{H}_1)$	$(-6.00\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.35\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.53\mathbf{E}^{-2},\mathbf{H}_1)$
10	*		*		*	$(-3.06\mathbf{E}^{-2},\mathbf{H}_0)$	$(-10.42\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.57\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.85\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.53\mathbf{E}^{-2},\mathbf{H}_1)$	$(-9.73\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.06\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.65\mathbf{E}^{-2},\mathbf{H}_0)$
11	*			*	*	$(-0.70\mathbf{E}^{-2},\mathbf{H}_0)$	$(-5.96\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.13\mathbf{E}^{-2},\mathbf{H}_0)$	$(+0.65\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.00\mathbf{E}^{-2},\mathbf{H}_0)$	$(-5.67\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.55\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.05\mathbf{E}^{-2},\mathbf{H}_0)$
12		*	*	*		$(-2.92\mathbf{E}^{-2},\mathbf{H}_1)$	$(-9.36\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.25\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.82\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.17\mathbf{E}^{-2},\mathbf{H}_1)$	$(-8.50\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.84\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.23\mathbf{E}^{-2},\mathbf{H}_1)$
13		*	*		*	$(-3.43\mathbf{E}^{-2},\mathbf{H}_1)$	$(-10.43\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.65\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.44\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.87\mathbf{E}^{-2},\mathbf{H}_1)$	$(-10.14\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.90\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.09\mathbf{E}^{-2},\mathbf{H}_1)$
14		*		*	*	$(-1.27\mathbf{E}^{-2},\mathbf{H}_0)$	$(-6.50\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.24\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.49\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.24\mathbf{E}^{-2},\mathbf{H}_0)$	$(-6.33\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.93\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.37\mathbf{E}^{-2},\mathbf{H}_0)$
15			*	*	*	$(-3.04\mathbf{E}^{-2},\mathbf{H}_0)$	$(-9.04\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.34\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.97\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.57\mathbf{E}^{-2},\mathbf{H}_1)$	$(-8.70\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.62\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.37\mathbf{E}^{-2},\mathbf{H}_1)$
16	*	*				$(-1.01\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.38\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.97\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.67\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.25\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.66\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.35\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.25\mathbf{E}^{-2},\mathbf{H}_0)$
17	*		*			$(-4.03\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.96\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.68\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.49\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.84\mathbf{E}^{-2},\mathbf{H}_1)$	$(-6.56\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.12\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.84\mathbf{E}^{-2},\mathbf{H}_1)$
18	*			*		$(-0.63\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.21\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.70\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.06\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.52\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.91\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.35\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.33\mathbf{E}^{-2},\mathbf{H}_0)$
19	*				*	$(-0.90\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.86\mathbf{E}^{-2},\mathbf{H}_0)$	$(-4.09\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.90\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.99\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.88\mathbf{E}^{-2},\mathbf{H}_0)$	$(-4.05\mathbf{E}^{-2},\mathbf{H}_1)$	$(-0.99\mathbf{E}^{-2},\mathbf{H}_0)$
20		*	*			$(-0.60\mathbf{E}^{-2},\mathbf{H}_0)$	$(-2.65\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.25\mathbf{E}^{-2},\mathbf{H}_0)$	$(-0.60\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.19\mathbf{E}^{-2},\mathbf{H}_0)$	$(-3.07\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.56\mathbf{E}^{-2},\mathbf{H}_0)$	$(-1.19\mathbf{E}^{-2},\mathbf{H}_0)$
21		*		*		$(-2.42\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.47\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.67\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.42\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.37\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.41\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.72\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.37\mathbf{E}^{-2},\mathbf{H}_1)$
22		*			*	$(-4.07\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.59\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.44\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.07\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.69\mathbf{E}^{-2},\mathbf{H}_1)$	$(-6.13\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.87\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.69\mathbf{E}^{-2},\mathbf{H}_1)$
23			*	*		$(-4.96\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.21\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.17\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.33\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.25\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.42\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.41\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.25\mathbf{E}^{-2},\mathbf{H}_1)$
24			*		*	$(-5.75\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.79\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.86\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.36\mathbf{E}^{-2},\mathbf{H}_1)$	$(-6.32\mathbf{E}^{-2},\mathbf{H}_1)$	$(-8.22\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.95\mathbf{E}^{-2},\mathbf{H}_1)$	$(-5.67\mathbf{E}^{-2},\mathbf{H}_1)$
25				*	*	$(-3.27\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.37\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.97\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.27\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.84\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.78\mathbf{E}^{-2},\mathbf{H}_1)$	$(-4.51\mathbf{E}^{-2},\mathbf{H}_1)$	$(-3.84\mathbf{E}^{-2},\mathbf{H}_1)$
26	-	-	-	-	-	$(-1.92\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.31\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.68\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.01\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.52\mathbf{E}^{-2},\mathbf{H}_1)$	$(-7.03\mathbf{E}^{-2},\mathbf{H}_1)$	$(-2.37\mathbf{E}^{-2},\mathbf{H}_1)$	$(-1.69\mathbf{E}^{-2},\mathbf{H}_1)$

Table 3: Wilcoxon tests comparing the performance of eight standard fusion approaches presented in [12].

	SIFT	SURF	ORB	BRISK	FREAK	t-norm $s = 0$	t-norm $s = 0.5$	t-norm $s = 1$
0	*	*	*	*	*	$(+3.79\mathbf{E}^{-2}, \mathbf{H}_1)$	$(-0.27\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+5.78\mathbf{E}^{-2}, \mathbf{H}_1)$
1	*	*	*	*		$(+5.70\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+0.13\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+5.27\mathbf{E}^{-2}, \mathbf{H}_1)$
2	*	*	*		*	$(+3.58\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+1.51\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+5.73\mathbf{E}^{-2}, \mathbf{H}_1)$
3	*	*		*	*	$(+5.10\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+2.14\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+7.17\mathbf{E}^{-2}, \mathbf{H}_1)$
4	*		*	*	*	$(+3.78\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+1.22\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+4.66\mathbf{E}^{-2}, \mathbf{H}_1)$
5		*	*	*	*	$(+3.11\mathbf{E}^{-2}, \mathbf{H}_1)$	$(-1.62\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+3.83\mathbf{E}^{-2}, \mathbf{H}_1)$
6	*	*	*			$(+4.42\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+2.12\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+5.12\mathbf{E}^{-2}, \mathbf{H}_1)$
7	*	*		*		$(+7.16\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.28\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+7.11\mathbf{E}^{-2}, \mathbf{H}_1)$
8	*	*			*	$(+4.79\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+5.04\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+7.00\mathbf{E}^{-2}, \mathbf{H}_1)$
9	*		*	*		$(+5.15\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+1.38\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+5.08\mathbf{E}^{-2}, \mathbf{H}_1)$
10	*		*		*	$(+2.47\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+2.10\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+4.37\mathbf{E}^{-2}, \mathbf{H}_1)$
11	*			*	*	$(+4.88\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+4.03\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+6.10\mathbf{E}^{-2}, \mathbf{H}_1)$
12		*	*	*		$(+1.15\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.97\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+1.70\mathbf{E}^{-2}, \mathbf{H}_1)$
13		*	*		*	$(+2.41\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+0.25\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+3.47\mathbf{E}^{-2}, \mathbf{H}_1)$
14		*		*	*	$(+4.42\mathbf{E}^{-2}, \mathbf{H}_1)$	$(-0.16\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+4.92\mathbf{E}^{-2}, \mathbf{H}_1)$
15			*	*	*	$(+2.67\mathbf{E}^{-2}, \mathbf{H}_1)$	$(-1.95\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+2.67\mathbf{E}^{-2}, \mathbf{H}_1)$
16	*	*				$(+6.14\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+5.92\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+6.54\mathbf{E}^{-2}, \mathbf{H}_1)$
17	*		*			$(+2.00\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+2.66\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+2.61\mathbf{E}^{-2}, \mathbf{H}_1)$
18	*			*		$(+6.67\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+7.07\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+7.19\mathbf{E}^{-2}, \mathbf{H}_1)$
19	*				*	$(+2.75\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.64\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.70\mathbf{E}^{-2}, \mathbf{H}_1)$
20		*	*			$(+3.32\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.29\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.69\mathbf{E}^{-2}, \mathbf{H}_1)$
21		*		*		$(+2.18\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+1.25\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+2.51\mathbf{E}^{-2}, \mathbf{H}_1)$
22		*			*	$(+3.36\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.50\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+4.09\mathbf{E}^{-2}, \mathbf{H}_1)$
23			*	*		$(+0.29\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.89\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.67\mathbf{E}^{-2}, \mathbf{H}_0)$
24			*		*	$(+1.89\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+0.47\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+1.90\mathbf{E}^{-2}, \mathbf{H}_1)$
25				*	*	$(+3.64\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+1.99\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+3.25\mathbf{E}^{-2}, \mathbf{H}_1)$
26	-	-	-	-	-	$(+3.72\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+1.81\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+4.47\mathbf{E}^{-2}, \mathbf{H}_1)$

Table 4: Wilcoxon tests comparing the performance of our fusion approach for different combination schemes.

	SIFT	SURF	ORB	BRISK	FREAK	2 vs 3	2 vs 4	3 vs 4
0	*	*	*	*	*	$(-0.38\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.49\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.12\mathbf{E}^{-2}, \mathbf{H}_0)$
1	*	*	*	*		$(+0.12\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.13\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$
2	*	*	*		*	$(-0.46\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.29\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.16\mathbf{E}^{-2}, \mathbf{H}_0)$
3	*	*		*	*	$(-0.44\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.12\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.32\mathbf{E}^{-2}, \mathbf{H}_0)$
4	*		*	*	*	$(-0.04\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.37\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.33\mathbf{E}^{-2}, \mathbf{H}_0)$
5		*	*	*	*	$(-0.08\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.19\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.26\mathbf{E}^{-2}, \mathbf{H}_0)$
6	*	*	*			$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.03\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.04\mathbf{E}^{-2}, \mathbf{H}_0)$
7	*	*		*		$(-0.61\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.60\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$
8	*	*			*	$(+0.76\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.87\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.11\mathbf{E}^{-2}, \mathbf{H}_0)$
9	*		*	*		$(+0.16\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.04\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.21\mathbf{E}^{-2}, \mathbf{H}_0)$
10	*		*		*	$(+0.15\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.17\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$
11	*			*	*	$(-0.07\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.77\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.70\mathbf{E}^{-2}, \mathbf{H}_0)$
12		*	*	*		$(+0.14\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.05\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.09\mathbf{E}^{-2}, \mathbf{H}_0)$
13		*	*		*	$(-0.20\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.16\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.03\mathbf{E}^{-2}, \mathbf{H}_0)$
14		*		*	*	$(-0.63\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.34\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.29\mathbf{E}^{-2}, \mathbf{H}_0)$
15			*	*	*	$(+0.00\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.10\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.10\mathbf{E}^{-2}, \mathbf{H}_0)$
16	*	*				$(-0.59\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.04\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.64\mathbf{E}^{-2}, \mathbf{H}_0)$
17	*		*			$(+0.32\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+1.06\mathbf{E}^{-2}, \mathbf{H}_1)$	$(+0.75\mathbf{E}^{-2}, \mathbf{H}_0)$
18	*			*		$(+0.81\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.89\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.07\mathbf{E}^{-2}, \mathbf{H}_0)$
19	*				*	$(+0.39\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.64\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.25\mathbf{E}^{-2}, \mathbf{H}_0)$
20		*	*			$(-0.01\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.23\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.22\mathbf{E}^{-2}, \mathbf{H}_0)$
21		*		*		$(-0.44\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.79\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.34\mathbf{E}^{-2}, \mathbf{H}_0)$
22		*			*	$(-0.14\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.02\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.16\mathbf{E}^{-2}, \mathbf{H}_0)$
23			*	*		$(-0.17\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.46\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.28\mathbf{E}^{-2}, \mathbf{H}_0)$
24			*		*	$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.4\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.42\mathbf{E}^{-2}, \mathbf{H}_0)$
25				*	*	$(-0.20\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.21\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.02\mathbf{E}^{-2}, \mathbf{H}_0)$
26	-	-	-	-	-	$(-0.06\mathbf{E}^{-2}, \mathbf{H}_0)$	$(-0.05\mathbf{E}^{-2}, \mathbf{H}_0)$	$(+0.01\mathbf{E}^{-2}, \mathbf{H}_0)$

Table 5: Wilcoxon tests comparing the performance of our fusion approach using different values of the parameter n. Results indicate that the value of n produce differences in performance that are not statistically significant (null hypothesis H_0 holds).

winner. So, we might conclude that the impact of the parameter n is not critical in the performance of the method. We believe it is because if the correct match is not amongst the first two, then, the descriptor is not reliable for that particular keypoint. In other words, it is unlikely for a descriptor to classify the correct match in the third or fourth position.

In order to analyze the behavior of our method as a function of β, we show the evolution of the F measure for different fusion combinations in Fig. 4 for a constant value of n and α. We have selected only a set of combinations since for the rest of cases the behavior is very similar. As can be observed, a smooth function is obtained. The values of F measure increase slowly until reaching its maximum and then the function decreases smoothly.

At the light of the results presented, it can be seen that the results of the proposed method are robust, i.e., they do not strongly depend on the parameter selection.

7. Conclusions

This work has proposed a new keypoint matching approach. Our method fuses two of more keypoint descriptors creating an evidence distribution. Then, an analysis of the entropy of the distribution is employed to calculate a confidence factor. The confidence factor is an indicator of how good a descriptor is at estimating the correct match of a keypoint. DST is employed to fuse the evidence distributions, and finally the pignistic probability is applied to select the most likely match.

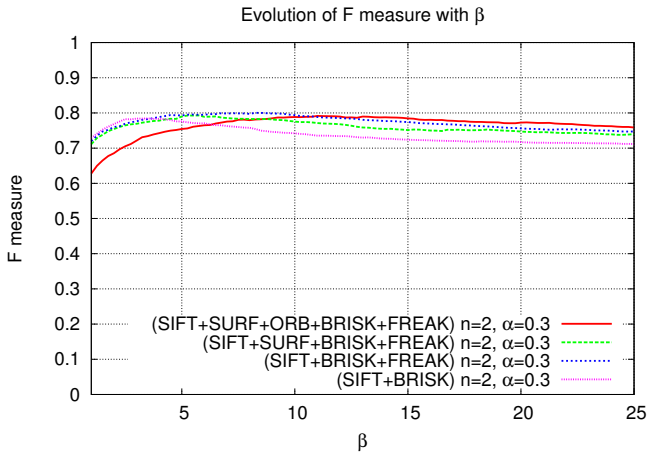

Figure 4: Evolution of the F measure as the smoothing parameter β changes (Eq.22). As can be observed, the performance of our method is very regular for a wide range of values of β.

The proposed method has been tested on the Oxford keypoint dataset [2] running statistical tests on the results. The tests conducted show that the proposed method obtains a performance improvement (which is statistically relevant) in the majority of the cases. Also, we have observed that the better the descriptors employed for fusion are, the higher the improvement obtained. In the best case, our method has obtained an improvement of 10% with respect to the best keypoint descriptor FREAK. Finally, our experiments have also shown that the performance of the proposed method is robust to a wide range of values of its parameters.

Acknowledgements

This work was partially supported by the Research Projects TIN2012-32952 and BROCA, both financed by the Spanish Ministry of Science and Technology and FEDER.

References

[1] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[2] C. Schmid, K. Mikolajczyk, A performance evaluation of local descriptors, in: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, Madison, Wisconsin, 2003, pp. 257–263. doi:10.1109/CVPR.2003.1211478.

[3] C. Schmid, R. Mohr, Local grayvalue invariants for image retrieval, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (5) (1997) 530–535. doi:10.1109/34.589215.

[4] M. Brown, D. G. Lowe, Automatic panoramic image stitching using invariant features, International Journal of Computer Vision 74 (1) (2007) 59–73. doi:10.1007/s11263-006-0002-3.

[5] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.

[6] D. Scharstein, R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision 47 (1-3) (2002) 7–42. doi:10.1023/A:1014573219977.

[7] J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation, Pattern Analysis and Machine Intelligence, IEEE Transactions on 25 (7) (2003) 787–800. doi:10.1109/TPAMI.2003.1206509.

[8] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded up robust features, in: Computer Vision ECCV 2006, Vol. 3951 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Graz, Austria, 2006, pp. 404–417. doi:10.1007/11744023_32.

[9] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, ORB: An efficient alternative to SIFT or SURF, in: Computer Vision (ICCV), 2011 IEEE International Conference on, Barcelona, Spain, 2011, pp. 2564–2571. doi:10.1109/ICCV.2011.6126544.

[10] S. Leutenegger, M. Chli, R. Y. Siegwart, BRISK: Binary robust invariant scalable keypoints, in: Computer Vision (ICCV), 2011 IEEE International Conference on, Barcelona, Spain, 2011, pp. 2548–2555. doi:10.1109/ICCV.2011.6126542.

[11] A. Alahi, R. Ortiz, P. Vandergheynst, FREAK: Fast retina keypoint, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, Rhode Island, Providence, USA, 2012, pp. 510–517. doi:10.1109/CVPR.2012.6247715.

[12] P. Perakis, T. Theoharis, I. A. Kakadiaris, Feature fusion for facial landmark detection, Pattern Recognition 47 (9) (2014) 2783–2793. doi:10.1016/j.patcog.2014.03.007.

[13] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, Pattern Analysis and Machine Intelligence, IEEE Transactions on 27 (10) (2005) 1615–1630. doi:10.1109/TPAMI.2005.188.

[14] E. Rosten, T. Drummond, Machine learning for high-speed corner detection, in: Computer Vision ECCV 2006, Vol. 3951 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Graz, Austria, 2006, pp. 430–443. doi:10.1007/11744023_34.

[15] M. Calonder, V. Lepetit, C. Strecha, P. Fua, BRIEF: binary robust independent elementary features, in: Computer Vision ECCV 2010, Vol. 6314 of Lecture Notes in Computer Science, Springer, Crete, Greece, 2010, pp. 778–792. doi:10.1007/978-3-642-15561-1_56.

[16] D. He, S. Liang, Y. Fang, A multi-descriptor, multi-nearest neighbor approach for image classification, in: Advanced Intelligent Computing Theories and Applications, International Conference on Intelligent Computing, ICIC 2010, Vol. 6215 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Changsha, China, 2010, pp. 515–523. doi:10.1007/978-3-642-14922-1_64.

[17] S. Bakshi, H. Mehrotra, R. Raman, P. K Sa, Score level fusion of SIFT and SURF for IRIS, in: Devices, Circuits and Systems (ICDCS), 2012 International Conference on, IEEE, Coimbatore, India, 2012, pp. 527 – 531. doi:10.1109/ICDCSyst.2012.6188740.

[18] P. Mountney, B. Lo, S. Thiemjarus, D. Stoyanov, G. Zhong-Yang, A probabilistic framework for tracking deformable soft tissue in minimally invasive surgery, in: Medical Image Computing and Computer-Assisted Intervention MICCAI 2007, Vol. 4792 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Brisbane, Australia, 2007, pp. 34–41. doi:10.1007/978-3-540-75759-7_5.

[19] R. Weng, J. Lu, J. Hu, G. Yang, Y.-P. Tan, Robust feature set matching for partial face recognition, in: Computer Vision (ICCV), 2013 IEEE International Conference on, Sydney, Australia, 2013, pp. 601–608. doi:10.1109/ICCV.2013.80.

[20] E. P. Xing, A. Y. Ng, M. I. Jordan, S. Russell, Distance metric learning with application to clustering with side-information, Advances in neural information processing systems (2002) 505–512.

[21] P. Smets, The combination of evidence in the transferable belief model, Pattern Analysis and Machine Intelligence, IEEE Transactions on 12 (1990) 447–458. doi:10.1109/34.55104.

[22] R. Munoz-Salinas, R. Medina-Carnicer, F. Madrid-Cuevas, A. Carmona-Poyato, Multi-camera people tracking using evidential filters, International Journal of Approximate Reasoning 50 (2009) 732–749. doi:10.1016/j.ijar.2009.02.001.

[23] S. Panigrahi, A. Kundu, S. Sural, A. K. Majumdar, Use of dempster-shafer theory and bayesian inferencing for fraud detection in mobile communication networks, in: Information Security and Privacy, Vol. 4586 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Townsville, Australia, 2007, pp. 446–460. doi:10.1007/978-3-540-73458-1_32.

[24] T. Denoeux, P. Smets, Classification using belief functions: the relationship between the case-based and model-based approaches, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 36 (2006) 1395–1406. doi:10.1109/TSMCB.2006.877795.

[25] S. Demotier, W. Schon, T. Denoeux, Risk assessment based on weak information using belief functions: A case study in water treatment, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 36 (2006) 382–396. doi:10.1109/TSMCC.2004.840057.

[26] T. Denoeux, M.-H. Masson, Evclus: Evidential clustering of proximity data, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 34 (2004) 95–109. doi:10.1109/TSMCB.2002.806496.

[27] M.-H. Masson, T. Denoeux, Ecm: An evidential version of the fuzzy c-means algorithm, Pattern Recognition 41 (4) (2008) 1384– 1397. doi:10.1016/j.patcog.2007.08.014.

[28] A.-O. Boudraa, A. Bentabet, F. Salzenstein, Dempster-shafer's basic probability assignment based on fuzzy membership functions, Electronic Letters on Computer Vision and Image Analysis 4 (2004) 1–10. doi:10.5565/rev/elcvia.68.

[29] I. Bloch, Defining belief functions using mathematical morphology - Application to image fusion under imprecision, International Journal of Approximate Reasoning 48 (2008) 437–465. doi:10.1016/j.ijar.2007.07.008.

[30] Z. Hammal, L. Couvreur, A. Caplier, M. Rombaut, Facial expression classification: An approach based on the fusion of facial deformations using the transferable belief model, International Journal of Approximate Reasoning 46 (2007) 542–567. doi:10.1016/j.ijar.2007.02.003.

[31] W. Pieczynski, Multisensor triplet markov chains and theory of evidence, International Journal of Approximate Reasoning 45 (2007) 1–16. doi:10.1016/j.ijar.2006.05.001.

[32] Z. Yi, H. Khing, C. Seng, Z. Wei, Multi-ultrasonic sensor fusion for autonomous mobile robots, in: SPIE proceedings series: Architectures, Algorithms and Applications IV, Orlando, Florida, 2000, pp. 314–321. doi:10.1117/12.381644.

[33] H. Wu, M. Siegel, R. Stiefelhagen, J. Yang, Sensor fusion using dempster-shafer theory, in: Instrumentation and Measurement Technology Conference, 2002. IMTC/2002. Proceedings of the 19th IEEE, Anchorage, AK, USA, 2002, pp. 7–12 vol.1. doi:10.1109/IMTC.2002.1006807.

[34] N. Milisavljevic, I. Bloch, S. Broek, M. Acheroy, Improving mine recognition through processing and dempster-shafer fusion of ground-penetrating radar data, Pattern Recognition 36 (2003) 1233–1250. doi:10.1016/S0031-3203(02)00251-0.

[35] P. Xu, F. Davoine, J.-B. Bordes, H. Zhao, T. Denoeux, Multimodal information fusion for urban scene understanding, Machine Vision and Applications (2014) 1–19doi:10.1007/s00138-014-0649-7.

[36] T. Denoeux, N. El Zoghby, V. Cherfaoui, A. Jouglet, Optimal object association in the dempster–shafer framework, Cybernetics, IEEE Transactions on 44 (12) (2014) 2521–2531. doi:10.1109/TCYB.2014.2309632.

[37] V.-N. Huynh, T. T. Nguyen, C. A. Le, Adaptively entropy-based weighting classifiers in combination using dempstershafer theory for word sense disambiguation, Computer Speech & Language 24 (3) (2010) 461 – 473. doi:10.1016/j.csl.2009.06.003.

[38] S. Ranoeliarivao, F. de Morsier, D. Tuia, S. Rakotoniaina, M. Borgeaud, J.-P. Thiran, S. Rakotondraompiana, Multisource clustering of remote sensing images with entropy-based dempster-shafer fusion, in: Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European, Marrakech, Morocco, 2013, pp. 1–5.

[39] D. Pan, X. Lu, J. Liu, Y. Deng, A ranking procedure by incomplete pairwise comparisons using information entropy and dempster-shafer evidence theory, The Scientific World Journal 2014. doi:10.1155/2014/904596.

[40] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence 172 (2–3) (2008) 234 – 264. doi:10.1016/j.artint.2007.05.008.

[41] R. Kennes, P. Smets, Computational aspects of the mobius transformation, in: Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence, UAI '90, Elsevier Science Inc., New York, NY, USA, 1991, pp. 401–416.

[42] B. Quost, M.-H. Masson, T. Denoeux, Classifier fusion in the dempster-shafer framework using optimized t-norm based combination rules, International Journal of Approximate Reasoning 52 (3) (2011) 353–374. doi:10.1016/j.ijar.2010.11.008.

[43] T. Denoeux, The cautious rule of combination for belief functions and some extensions, in: Information Fusion, 9th International Conference on, Florence, Italy, 2006, pp. 1–8. doi:10.1109/ICIF.2006.301572.

[44] P. Smets, Belief functions: The disjunctive rule of combination and the generalized bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35. doi:10.1016/0888-613X(93)90005-X.

[45] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (2) (1994) 191–243. doi:10.1016/0004-3702(94)90026-4.

[46] F. Cuzzolin, The intersection probability and its properties, in: Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Vol. 5590 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Verona, Italy, 2009, pp. 287–298. doi:10.1007/978-3-642-02906-6_26.

[47] F. Cuzzolin, Geometry of relative plausibility and relative belief of singletons, Annals of Mathematics and Artificial Intelligence 59 (2010) 47–79. doi:10.1007/s10472-010-9186-x.

[48] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) (1945) 80–83.