Machine Vision and Applications (2014) 25:763–773
DOI 10.1007/s00138-014-0593-6

ORIGINAL PAPER

ReigSAC: fast discrimination of spurious keypoint correspondences on planar surfaces

Hugo Proença

Received: 16 November 2012 / Revised: 3 January 2014 / Accepted: 6 January 2014 / Published online: 7 February 2014

Abstract Various methods were proposed to detect/match special interest points (keypoints) in images and some of them (e.g., SIFT and SURF) are among the most cited techniques in computer vision research. This paper describes an algorithm to discriminate between *genuine* and *spurious* keypoint correspondences on planar surfaces. We draw random samples of the set of correspondences, from which homographies are obtained and their principal eigenvectors extracted. Density estimation on that feature space determines the most likely *true* transform. Such homography feeds a cost function that gives the goodness of each keypoint correspondence. Being similar to the well-known RANSAC strategy, the key finding is that the main eigenvector of the most (genuine) homographies tends to represent a similar direction. Hence, density estimation in the *eigenspace* dramatically reduces the number of transforms actually evaluated to obtain reliable estimations. Our experiments were performed on hard image data sets, and pointed that the proposed approach yields effectiveness similar to the RANSAC strategy, at significantly lower computational burden, in terms of the proportion between the number of homographies generated and those that are actually evaluated.

Keywords Keypoints detection · Keypoints matching · RANSAC

1 Introduction

The use of local descriptors in various computer vision applications has been gaining popularity. There are several justifications for using such techniques: (1)descriptors are distinctive; (2) robust to occlusion; and (3) do not require accurate data segmentation and normalization. This kind of techniques proved their effectiveness in a broad range of applications, including object recognition (e.g., [15,22] and [24]), image alignment (e.g., [17]), tracking (e.g., [9]) and robot navigation (e.g., [13]).

The value given to descriptors corresponds to their invariance in transformative classes, including translation, scale, rotation, affine and perspective changes. Various description and matching techniques were proposed, based on color, intensity, texture or gradient information [1,3,18] and [28]. This kind of methods can be classified according to two criteria: (1) their invariance levels; and (2) the analyzed data. Empirical evaluations of such genre of techniques can be found in the literature [20,25] and [26].

Regardless of the effectiveness of keypoint detection and matching techniques, the occurrence of false correspondences between keypoints is unavoidable and leads to errors in subsequent processing phases. Hence, strategies to discriminate between the genuine and spurious keypoint correspondences are of interest. Among others, a family of algorithms based on the Random Sample Consensus (RANSAC) [10] gained popularity, due to its simple implementation and robustness. It is divided into two main phases: (1) generation of a set of hypothesis from random samples; and (2) validation of each hypothesis with respect to the remaining data. Its use is widely reported in the literature, in domains such as model fitting, epipolar geometry estimation, and feature-based localization. However, the disadvantage frequently reported is the high-computational burden in case that too many iterations are required for convergence, which happens for high order geometric hypothesis and heavy contaminated data. This observation is the key motivation behind the work described in this paper.

H. Proença (✉)
Department of Computer Science, IT-Instituto de Telecomunicacoes, University of Beira Interior, 6200 Covilha, Portugal
e-mail: hugomcp@di.ubi.pt

1.1 Related work

A method similar to the proposed in this paper is due to Goshen and Shimshoni [12]. Here, authors described a method to automatically estimate the epipolar geometry in cases where correspondences are contaminated by many spurious cases (outliers). Their strategy is based on simple *Weak Motion Models*, each one sustained in an affine transform. Next, the inliers and outliers probabilities for the correspondences are obtained in a RANSAC-like way, based on corners statistics on both images.

The analysis of the state-of-the-art stresses the role played by RANSAC-like algorithms and their variants, which can be divided into attempts to improve: (1) accuracy; and (2) performance. Torr and Zisserman [30] presented a generalization of the RANSAC estimator (MLESAC) that shares the insight of RANSAC, but chooses the solution that maximizes the likelihood, instead of the one that maximizes the number of inliers. The error is modeled as a mixture of Gaussians and is assumed that the distribution of outliers is uniform, enabling the error to be minimized by the negative log likelihood. Given an initial estimate, authors suggest the Expectation-Maximization algorithm to estimate the parameters of the mixture model. Chum and Matas [6] used the ideas of order rank and of feature dissimilarity to propose a method to speed-up RANSAC. The rationale is to sort the set of correspondences based on the similarity function that attempts to establish correspondences. Then, assuming that such similarity measure is a weak predictor of genuine correspondences, they drew random samples from progressively larger sets of top-ranked correspondences. Authors observed substantial improvements in performance, and in the worst case, a convergence toward the RANSAC method. A similar idea was due to Tordoff and Murray [29], which used the prior probabilities of the validities between correspondences. These were observed to give an order of magnitude speed increase for problems where the correspondences are described by one image transformation and clutter. Additionally, authors showed how all putative correspondences, rather than just the best, from a particular feature can be taken forward into the sampling stage. Albeit at the expense of additional computation, this modification makes the algorithm suitable to handle correspondence sets modeled by two transformations and clutter at video-rates. Also with performance concerns, Ni et al. [21] speeded up the convergence process, by analyzing the group information between candidates provided by optical flow-based clustering. The assumption is that natural grouping criteria exist among the candidates, and used a binomial mixture model, particularly efficient in sampling data by reducing the number of groups from where candidates are selected and the number of candidates per group. Chin et al. [4] were also motivated by the computational burden of RANSAC, and proposed an approach to accelerate hypothesis sampling, using information derived from residual sorting, encoded in a series of inlier probabilities estimates. Authors observed that the proposed strategy dramatically reduces the number of samples required to converge to a satisfactory solution and, in comparison with other performance-concerned techniques, achieves best accuracy. Nistèr [23] proposed a variant of RANSAC for estimation of structure and motion in real-time environments, the main singularity of which is that instead of counting the number of potential inliers, the used scoring function is a robust likelihood or Bayesian cost function. Also, authors used preemptive scoring of the hypotheses generated by random sampling, to avoid excessive scoring of useless hypotheses contaminated by outliers or distorted by noise. Even considering that this might minimize the confidence in the selected hypothesis, the real-time requirements led to that choice. Kenney et al. [16] proposed a method based on the minimization of local objective functions for translation, rotation-scaling-translation and affine transforms. Shi and Tomasi [27] analyzed the eigenvalues of the normal matrix and presented a technique for measuring the quality of local image patches with respect to translation or affine transforms.

A comprehensive summary and performance evaluation of RANSAC-based techniques can be found in the work of Choi et al. [5]. Here, authors analyzed the most relevant techniques published in terms of accuracy, performance and robustness criteria.

1.2 Proposed method: cohesive perspective

When compared to the RANSAC algorithm, the method proposed in this paper has the following singularities: (1) it dramatically reduces the number of hypothesis generated for convergence, significantly reducing the computational burden; (2) it has a much more limited scope, being suitable exclusively for the purpose of estimate the goodness of keypoint correspondence on roughly planar data. The key steps are the use of density estimation techniques in the eigenspace of hypothesis (homographies), which reduces dramatically the number of transforms actually evaluated, and significantly lowers the computational burden of the process. In terms of spectral theory, eigenvectors can be regarded as invariant directions under a linear transformation. In this case, the main eigenvector can be regarded as the major distortion in the transform (homography) and giving an idea of the main orientation behind that distortion. The main eigenvectors of the homography matrices have only two degrees-of-freedom and for most (genuine) homographies, the main direction between the associated linear transform tends to be the approximately the same. This is the main key insight into the proposed algorithm.

Also, it should be stated the given method does not assume any constraint on the spatial relationship of genuine corre-

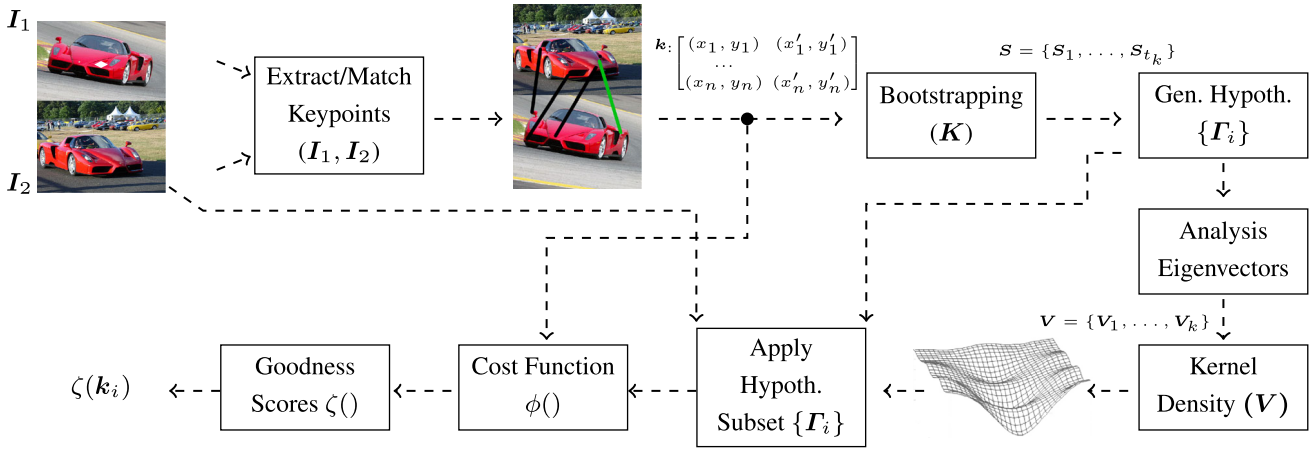

Fig. 1 Cohesive perspective of the proposed method. From a set of n keypoint correspondences ($\boldsymbol{K}$), a bootstrapping-based technique draws t_k random samples. Then, a set of hypothesis (homographies) $\{\boldsymbol{\Gamma}_i\}$ is found, and the largest eigenvectors of the matrix corresponding to their coordinates are projected into a feature space. Using kernel-based density estimation techniques, regions of maximal density are used to generate a score that outputs a goodness of fit value for each correspondence. Finally, a sigmoid function $\zeta(\boldsymbol{k}_i)$ gives the likelihood score of a correspondence being *genuine*

spondences and can be applied to any planar object, which is in opposition to previous studies that rely on geometrical constraints specific to each application domain (e.g., [8]). Figure 1 provides a cohesive perspective: the key insight is that genuine keypoint correspondences should come together in homographies that map the sample into the model data. We iteratively draw random samples from the set of correspondences and estimate the homography that match the model and the sample data. Their principal eigenvectors are projected into a 3D feature space, which enables locating the sub-space of maximal density. Then, a cost function based on the ℓ-2 norm estimates the goodness of each keypoint correspondence.

The remainder of this paper is organized as follows. Section 2 details each phase of the proposed method. Section 3 describes the used data sets and discusses the results obtained by our method. Section 4 presents our overall conclusions.

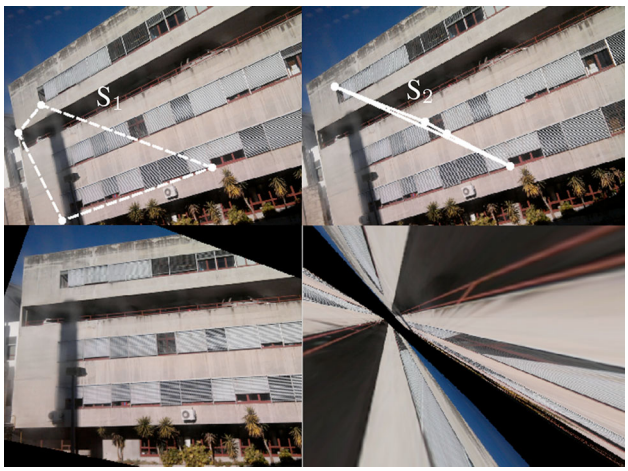

Fig. 2 Example of a homography extracted from two samples $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$. The sample in the top left image satisfies (1), in opposition to the sample given in the top right. Images at the bottom give the corresponding projected data, in which the degenerative case of the right image is evident because of the near-collinearity of three points in $\boldsymbol{S}_2$

2 Proposed method

2.1 Bootstrapping candidate selection

At least four point correspondences $\boldsymbol{k}_i$ are required to estimate a homography that maps points in image $\boldsymbol{I}_1$ into $\boldsymbol{I}_2$. Let $\boldsymbol{K} = [\boldsymbol{k}_1, \ldots, \boldsymbol{k}_n]$ be the set of keypoint correspondences between $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$, such that $\boldsymbol{k}_i = [x_i, y_i, x'_i, y'_i]$ specify the location (column x and row y) of the corresponding points on both images. Our rationale is to draw a reasonably large number of samples from $\boldsymbol{K}$ and estimate one homography per sample. Each sample $\boldsymbol{S}_i = \{s_1, \ldots, s_4\}$, being each s_i an index in $\{1, \ldots, n\}$ (sampling without replacement) that refers to rows of $\boldsymbol{K}$.

Let $\boldsymbol{C}$ be the set of possible combinations of $\boldsymbol{S}_i$, taken three at a time. A bootstrap sample $\boldsymbol{S}_i$ is considered valid if:

$$\forall c_i \in C, \left| det \begin{bmatrix} x_{c_{i1}} & y_{c_{i1}} & 1 \\ x_{c_{i2}} & y_{c_{i2}} & 1 \\ x_{c_{i3}} & y_{c_{i3}} & 1 \end{bmatrix} \right| > \delta, \tag{1}$$

where $|.|$ is the absolute value and $det[.]$ is the matrix determinant. In our experiments, we used $\delta = 0.05$, with normalized coordinates ($0 \leq (x, y) \leq 1$). The purpose of (1) is to assure that no three keypoints from any sample are collinear, to avoid degenerating homographies, as illustrated in Fig. 2.

2.2 Homography

When imaging by a perspective camera, any two images on the same planar surface are related by a homography $\boldsymbol{\Gamma}$ (also called a projective transform or collineation). Formally, it is given by:

$$[x, y, z]^T = \boldsymbol{\Gamma} \cdot [x'y'z']^T = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}, \tag{2}$$

where (x, y, z) and (x', y', z') are the homogenous coordinates in $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$. Setting $\{z, z'\} = 1$ yields a plane-to-plane homography:

$$\begin{cases} x(h_{31}x' + h_{32}y' + h_{33}) = h_{11}x' + h_{12}y' + h_{13} \\ y(h_{31}x' + h_{32}y' + h_{33}) = h_{21}x' + h_{22}y' + h_{23}. \end{cases} \tag{3}$$

Each keypoint correspondence relates a point in $\boldsymbol{I}_1$ to a point in $\boldsymbol{I}_2$. Given a set of correspondences, a matrix $\boldsymbol{A}$ can be obtained by concatenating the Eq. (3), resulting in a system of linear equations:

$$\boldsymbol{A} \cdot \boldsymbol{\Gamma} = 0, \tag{4}$$

that can be solved for $\boldsymbol{\Gamma}$ using homogenous linear least squares. Hartley and Zisserman [14] give additional details about this kind of techniques.

2.3 Eigenvector analysis

Let $\boldsymbol{\Gamma} = \{p_{ij}\}$, $1 \leq i, j \leq 3$ be a homography. $\boldsymbol{I}_3$ is the identity matrix and $\boldsymbol{C}$ a scalar matrix, such that $\boldsymbol{C} = \lambda \boldsymbol{I}_3$.

$$\boldsymbol{\Gamma} - \boldsymbol{C} = \begin{bmatrix} p_{11} - \lambda & p_{12} & p_{13} \\ p_{21} & p_{22} - \lambda & p_{23} \\ p_{31} & p_{32} & p_{33} - \lambda \end{bmatrix} \tag{5}$$

$det(\boldsymbol{\Gamma} - \boldsymbol{C}) = 0$ is a cubic equation which roots yield the eigenvalues $(\lambda^{(i)}, i = 1, 2, 3)$ of $\boldsymbol{\Gamma}$. For each eigenvalue, the corresponding eigenvector $\boldsymbol{v}^{(i)}$ is given by:

$$(\boldsymbol{\Gamma} - \lambda^{(i)})\boldsymbol{v}^{(i)} = 0. \tag{6}$$

Let $\boldsymbol{V} = [\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \boldsymbol{v}^{(3)}]$, ($\boldsymbol{v}^{(i)}$ are column vectors) be the eigenvectors matrix, which might have complex components. The Schur [11] decomposition was used to convert it into the real block diagonal form $\boldsymbol{U}$. Because $\boldsymbol{U}$ is similar to $\boldsymbol{\Gamma}$, it shares the multiset of eigenvalues. In the following, $\boldsymbol{v}$ denotes the principal eigenvector of $\boldsymbol{\Gamma}$.

2.4 Density estimation

The next step is to estimate the density in the above described feature space of $\boldsymbol{v}$ elements, which was done according to a kernel-based strategy, as described by Matej et al. [19]. Their model $p_s(\boldsymbol{v})$ of d-dimensional data is given by a Gaussian mixture model with t_g components:

$$p_s(\boldsymbol{v}) = \sum_{i=1}^{t_g} \alpha_i \frac{e^{-1/2(\boldsymbol{v}-\mu)^T \Sigma^{-1}(\boldsymbol{v}-\mu)}}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}}, \tag{7}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ denote the mean and covariance matrix of a Gaussian kernel. Considering $p_s(\boldsymbol{v})$ to be a sample distribution, the density estimate is given by its convolution with a kernel of covariance matrix $\boldsymbol{H}$ that defines bandwidth.

$$\hat{p}_{kde}(\boldsymbol{v}) = \sum_{i=1}^{t_g} \alpha_i \frac{e^{-1/2(\boldsymbol{v}-\boldsymbol{v}_i)^T (\boldsymbol{H}+\boldsymbol{\Sigma}_i)^{-1}(\boldsymbol{v}-\boldsymbol{v}_i)}}{(2\pi)^{d/2}|\boldsymbol{H} + \boldsymbol{\Sigma}|^{1/2}}, \tag{8}$$

where $\boldsymbol{v}_i$ specifies the center of the ith kernel. To find the kernel bandwidth $\boldsymbol{H}$ that minimizes the distance between the estimated model $\hat{p}_{kde}(\boldsymbol{v})$ and a probability density function, authors used the empirically observed covariance of the samples to approximate an error measure for the estimator $\hat{p}$ (asymptotic mean integrated squared error). Further details about the algorithm to obtain the density estimates can be found in [19].

2.5 Cost function

According to the procedure described in Sect. 2.4, we obtain an estimate of the densities in the feature space populated by the main eigenvector of each generated homography. Next, we find the maximum density point and define a neighborhood around it, selecting a subset of these homographies and obtaining a *cost* for each one. The cost function ϕ of each homography $\boldsymbol{\Gamma}$ is given by:

$$\phi(\boldsymbol{\Gamma}) = \frac{1}{t_k}\Big(\sum_{i=1}^{t_k} ||(x_i, y_i) - \boldsymbol{\Gamma}(x'_i, y'_i)||_2 + ||(x'_i, y'_i) - \boldsymbol{\Gamma}^{-1}(x_i, y_i)||_2\Big), \tag{9}$$

where $\boldsymbol{\Gamma}(x'_i, y'_i)$ denotes the location of the ith keypoint on the sample image transformed by $\boldsymbol{\Gamma}$, and t_k is the number of keypoint correspondences considered. It should be stressed that is not advisable to use all keypoint correspondences to evaluate a homography ($t_k = n$), as some correspondences will be unavoidably spurious. Instead, we observed that by considering a relatively short number of correspondences ($t_k \in [\frac{n}{3}, \frac{n}{2}]$), the best performance is obtained. As this parameter was empirically observed not to determine the over-

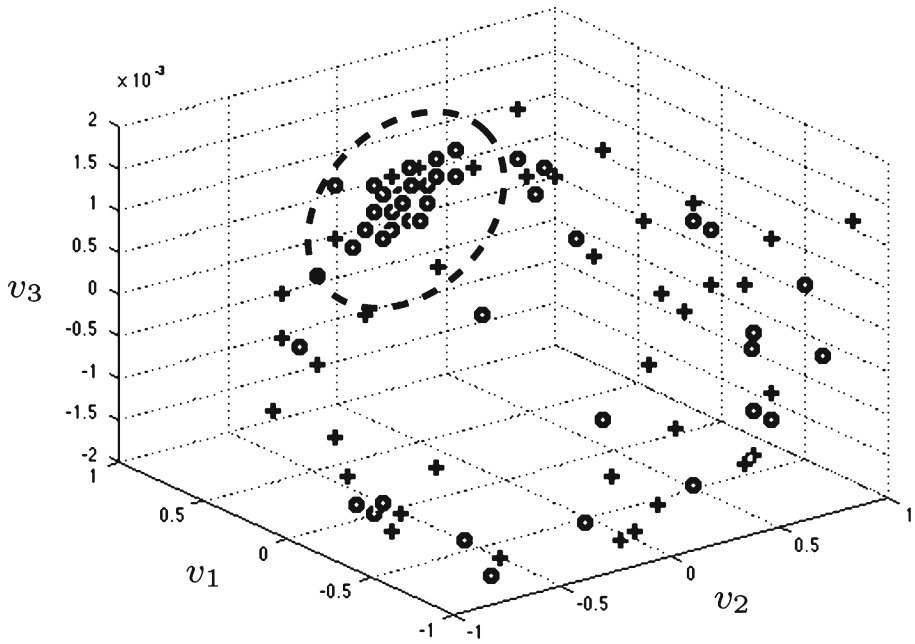

Fig. 3 Key insight of the proposed method: the feature space populated by the principal eigenvectors $\boldsymbol{v}$ of the homography matrices found from samples. The circular data points denote elements extracted exclusively from *genuine* keypoint correspondences, whereas the cross data points resulted from samples containing at least one *spurious* correspondence. In the region of highest density, a majority of genuine correspondences is evident, in contrast with the eigenvectors extracted from spurious correspondences that spread more across the feature space

all effectiveness of the proposed method in a strong way, this *optimal interval* was found to be adequate, instead of obtaining a unique value that will be strongly data-dependent. Also, it should be stressed that not all homographies are evaluated. As illustrated in Fig. 3, only those that are in the ℓ-2 norm neighborhood around the maximum density point in the feature space from $p_{kde}(\boldsymbol{v})$ are actually evaluated ($\approx$ 10 % below the maximum density was used in our experiments).

2.6 Reliability score

Finally, an estimate of the goodness of each keypoint correspondence is found, according to values obtained in (9). Let $m = \arg_l \min \phi(\boldsymbol{\Gamma}_l)$ be the index of the homography that minimizes the cost of projecting keypoints between the pair of images. The likelihood that a keypoint correspondence $\boldsymbol{k}_i$ is genuine has inverse correspondence to:

$$\zeta(\boldsymbol{k}_i) = ||(x_i, y_i) - \boldsymbol{\Gamma}_m(x_{i'}, y_{i'})||_2 + ||(x_{i'}, y_{i'}) - \boldsymbol{\Gamma}_m^{-1}(x_i, y_i)||_2 \quad (10)$$

3 Performance evaluation

The proposed method was mainly tested with two keypoint detection and matching techniques that are widely seen in the specialized literature: SIFT and SURF. The idea of both methods is to find positions in the images with notorious differences in appearances with respect to their neighbourhood (in space and scale terms). Further details about these techniques can be found at [18] and [1].

Proposed by Lowe [18], SIFT is an extremely popular technique for detecting and matching image keypoints. Detection starts by constructing a scale-space pyramid, approximated by the difference between images blurred by Gaussian kernels of different sigma values. Each pixel is considered a candidate if it is a local maximum within a 3D neighbourhood window. Next, keypoints with unstable extrema and corresponding to edge responses are rejected and the magnitude of the remaining keypoints is given by the accumulated energy of forward differences.

3.1 Data sets

For reproducibility, a concern was to keep all data sets used in the empirical evaluation of the proposed method publicly available[1]. The method was tested in a set of 1,000 images, acquired by a *Canon Digital Ixus 99 IS* camera. Images have originally dimensions 4,000 × 3,000 and were resized to 800 × 600 using bilinear interpolation techniques, converted to grayscale and stored in a bitmap format. To assure that a single homography can map the model into the sample data for every pair of images, all acquired scenes are roughly planar. All contain non-deformable objects acquired from various distances, 3D angles and non-controlled lighting conditions. Images contain both natural and human-made scenes as well as indoor and outdoor and paintings. Figure 4 presents several examples of the data used.

[1] http://www.di.ubi.pt/~hugomcp/doc/keyMatchData.zip.

Fig. 4 Examples of the data sets used in our experiments. The images regard a broad spectrum of scenes (e.g., buildings, vehicles, forests, paintings) and were acquired in non-controlled conditions with significant variations in the traditional data variation factors: indoor and outdoor scenes with scale, rotations, perspectives, occlusions, reflections, lighting and shadow variations

Ground-truth data are also available. For each pair of images with regard to the same scene, the SIFT and SURF keypoint detections and matching techniques were used and their outputs manually validated by human observation. Each keypoint correspondence found was binary classified into *genuine* or *spurious*, whether they regarded the same point on both the model and sample data.

Additionally, to provide a comparison term with related approaches, the proposed method was tested in the planar subset of the Oxford VGG data[2]. Thus, results for *grafitti*, *wall*, and *bark* images are also given.

3.2 Results and discussion

Figure 5 illustrates several examples of the proposed method, in which the keypoint correspondences deemed *genuine* are represented as *thin green lines* and the correspondences deemed *spurious* are represented as *thick red lines*. Here, *TPR* measures the system sensitivity, i.e., the proportion of genuine correspondences that were found genuine by the algorithm. Oppositely, the *FPR* value quantifies the proportion of spurious correspondences that were considered genuine by our algorithm.

Figure 6 gives a comparison between the probability density functions of $\zeta(\boldsymbol{k}_i)$ values for genuine and spurious keypoint correspondences, when using the SIFT (*top row*) and SURF (*bottom row*) detection and matching techniques. The *dashed lines* represent the genuine correspondences and the *continuous lines* represent the spurious correspondences. $\zeta(\boldsymbol{k}_i)$ are the raw score values and $\zeta^*(\boldsymbol{k}_i)$ are the transformed values that approximate Gaussian distributions, using the Box-Cox transform. The larger plots given at the right side compare both distributions (genuine/spurious correspondences), giving broad evidence of a clear separability between the typical $\zeta(\boldsymbol{k}_i)$ values observed for both types of keypoints comparisons, which is the root of the method proposed in this paper.

Let S be a random variable that represents the number of spurious correspondences selected when drawing a sample of cardinality t_k from the set of n keypoint correspondences, with s being spurious. The probability of selecting at least one spurious correspondence is given by:

$$\begin{aligned} P(S>0) &= 1 - P(S=0) \\ &= 1 - \frac{\binom{n-s}{t_k}}{\binom{n}{t_k}} \end{aligned} \tag{11}$$

In our experiments, the number of samples generated by the bootstrapping strategy was defined according to a user-adjusted reliability score. Let $P_i(S=0)$ be the probability of drawing a sample composed exclusively of genuine correspondences at iteration i. The corresponding probability mass function is given by:

$$P_i(S=0) = \Big(1 - P(S=0)\Big)^{i-1} P(S=0) \tag{12}$$

Accordingly, the cumulative distribution function is given by:

$$P_{<i}(S=0) = \sum_{k_i=1}^{i} \Big(1 - P(S=0)\Big)^{k_i-1} P(S=0) \tag{13}$$

The number of samples generated in the bootstrapping strategy is given by:

$$\arg\min_i P_{<i}(S=0) > \varsigma, \tag{14}$$

where ς is the reliability score (≈ 0.99 was used in our experiments).

To assess performance decreases regarding the *reliability* factor, we introduced spurious correspondences in the data and observed the resulting performance. However, as illustrated in the leftmost plots in Fig. 6, the densities for the $\zeta^*(\boldsymbol{k}_i)$ scores both for genuine and spurious keypoint correspondences are not Gaussian. Hence, before obtaining the decidability index, the $\zeta^*(\boldsymbol{k}_i)$ scores were firstly transformed to be roughly Gaussian, using the Box-Cox transformation [2]:

$$\zeta^*(\boldsymbol{k}_i) = \begin{cases} \frac{\zeta(\boldsymbol{k}_i)-1}{\lambda}, & \text{if } \lambda \neq 0 \\ log(\zeta(\boldsymbol{k}_i)), & \text{if } \lambda = 0 \end{cases}. \tag{15}$$

[2] http://www.robots.ox.ac.uk/~vgg/research/affine/.

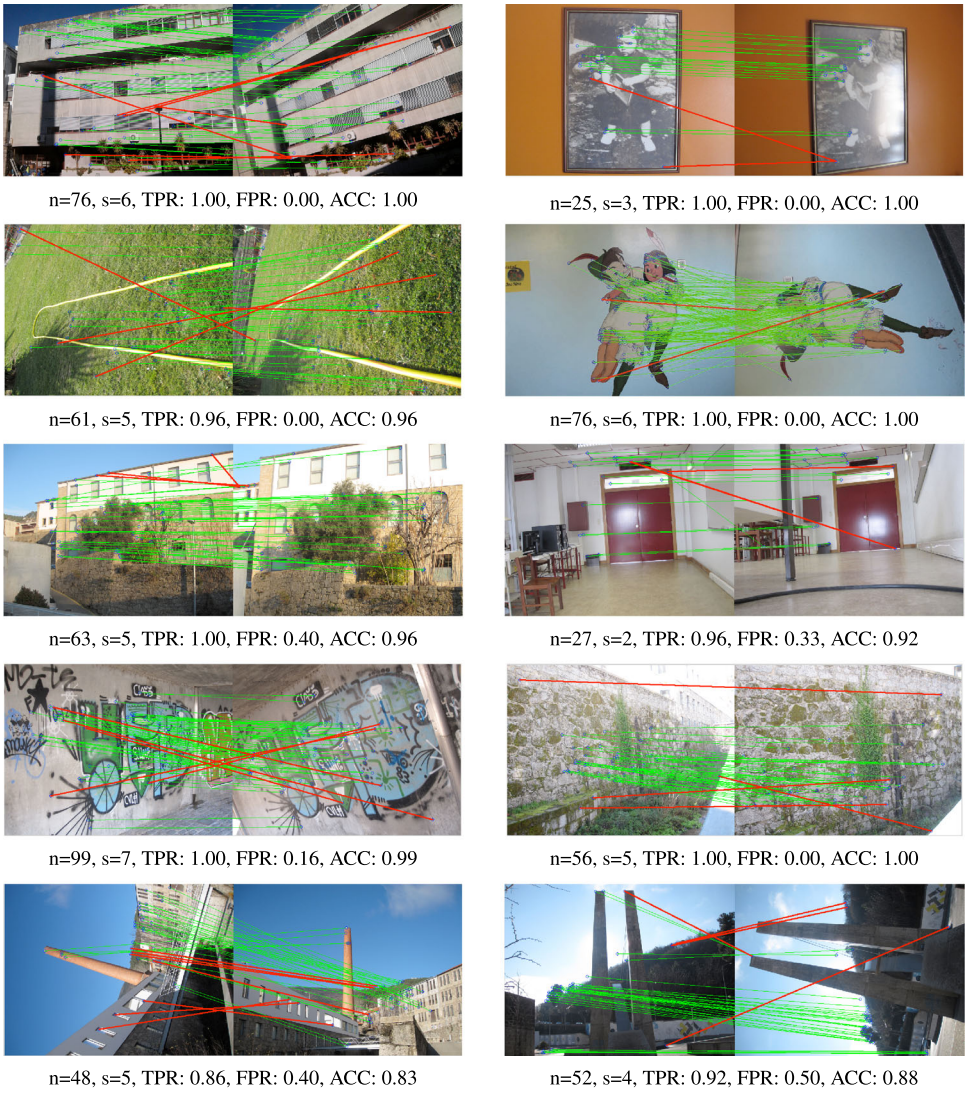

Fig. 5 Examples of the results obtained by the proposed method, in discriminating between the genuine (*thin green lines*) and spurious (*thick red lines*) keypoint correspondences. *TPR* sensitivity, *FPR* false-positive rate and *ACC* accuracy. *n* and *s* are the total of correspondences and the number of *spurious*. Images ate the bottom row illustrate not so successful cases: in these cases, the depths in scenes have strong variations and—as such—occurring transformations in data cannot be modeled by a single homography

Having a set of quasi-normal $\zeta^*(\boldsymbol{k}_i)$ scores, the decidability index was used to measure the average separation between the scores generated for genuine and spurious keypoint correspondences:

$$d' = \frac{|\mu_G - \mu_S|}{\sqrt{\frac{\sigma_G^2+\sigma_S^2}{2}}}, \tag{16}$$

where $\mu_G = \frac{\sum_i \zeta_i^{*G}}{k}$ and $\mu_S = \frac{\sum_i \zeta_i^{*S}}{m}$ are the means of the genuine/spurious distributions and $\sigma_G = \frac{\sum_i (\zeta_i^{*G}-\mu_G)^2}{k-1}$ and $\sigma_S = \frac{\sum_i (\zeta_i^{*S}-\mu_S)^2}{m-1}$ are their standard deviations.

Figure 7 presents two plots related to the issue of the proportion of spurious correspondences and the number of transforms (homographies) actually generated and evaluated with respect to that parameter: (1) the upper plot is the probability of selecting no spurious correspondences from a set of n correspondences with respect to the number of spurious correspondences (s); and (2) the bottom plot gives the relationship between the prior probability of drawing a sample composed exclusively by genuine correspondences ($P(S = 0)$), the number of homographies generated (light boxes) and those that were actually evaluated by (9) (dark boxes).

3.3 Computational complexity and results variability

The computational complexity of the proposed method is a major issue for real-time data handling. For comprehensibility, we denote every parameter evolved in the computational cost of the proposed method by $\alpha_.$ and a corresponding index (Table 1). The process starts by generating a random permutation, which can be performed in time linear to the number of keypoint correspondences $O(\alpha_1)$. Obtaining each homography involves computing the null space of the correspondences matrix (4), which takes $O(\alpha_2^2.\alpha_3)$. Each set of eigenvectors and eigenvalues was obtained in time $O(\alpha_4^4)$,

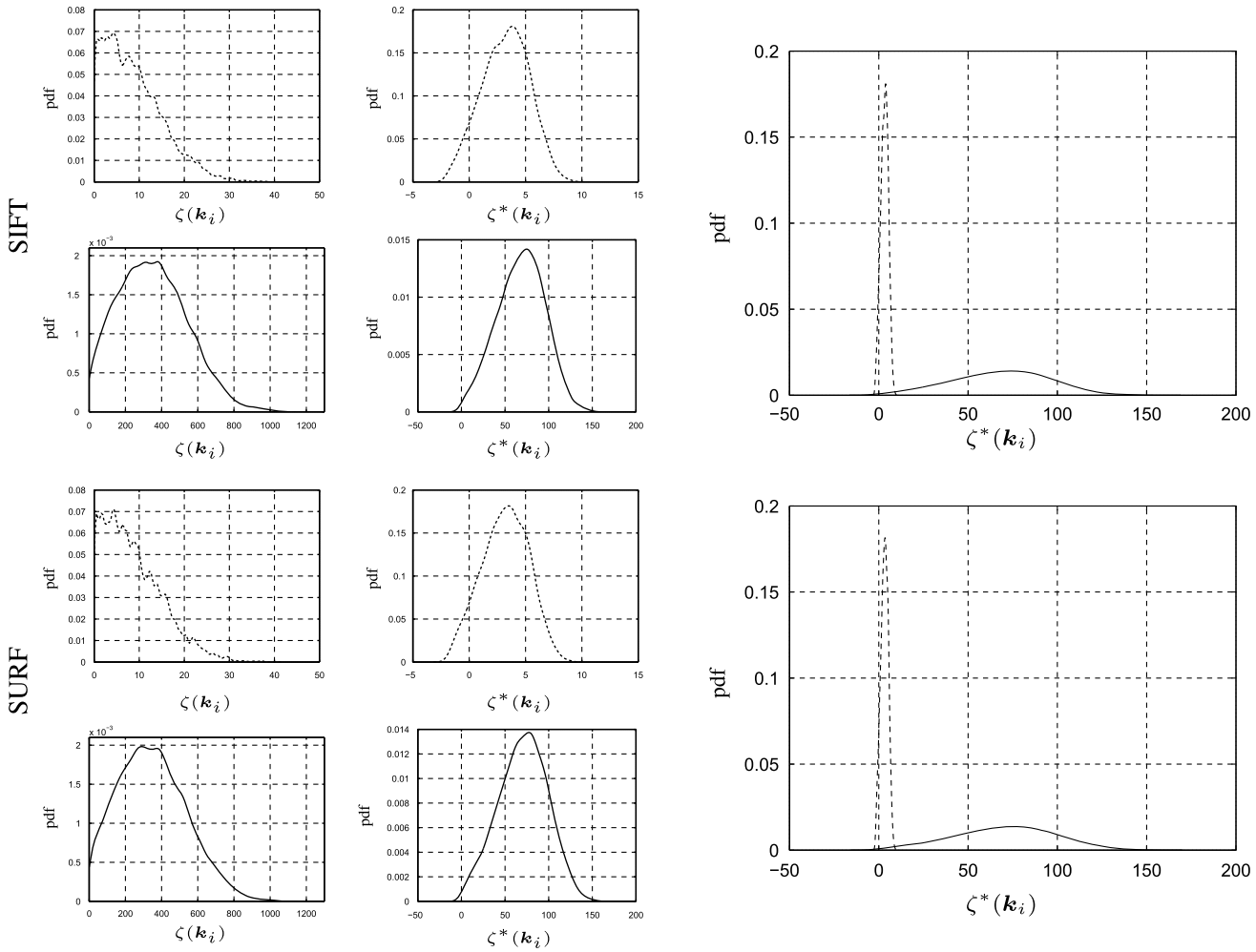

Fig. 6 Comparison between the probability density functions of $\zeta(\boldsymbol{k}_i)$ values for genuine and spurious keypoint correspondences, when using the SIFT (*top row*) and SURF (*bottom row*) detection and matching techniques. The *dashed lines* represent the genuine correspondences and the *continuous lines* represent the spurious correspondences. $\zeta(\boldsymbol{k}_i)$ are the raw score values and $\zeta^*(\boldsymbol{k}_i)$ are the transformed values that approximate Gaussian distributions, using the Box-Cox transform

where α_4 is the dimensions of the square matrix (three in our experiments). The transformation of sample data according to a homography is accomplished by matrix multiplication, performed in time $O(\alpha_5.\alpha_4^2)$, where α_5 is the number of keypoint correspondences.

In terms of the computational performance of the proposed algorithm, even though it is a data-dependent stochastic process, the most important factor is given in Fig. 7, plotting the proportion between the number of homographies generated (light boxes) and those that were actually evaluated by (9) (dark boxes). The major point is that a roughly quadratic correspondence was observed for both values, keeping values that are notoriously lower than that performed by the RANSAC algorithm (around 1,600 for $P(S = 0) = 0.1$) (Fig. 8).

As this procedure can be classified as a stochastic process, another issue concerns the variability of the results for a given input. Figure 9 gives the variability of the decidability scores (16) obtained, with respect to the reliability score ς. The results are expressed by boxplots, showing the median of the observed performance range (horizontal solid line) and the first and third quartile observation values (top and bottom of the box marks). The horizontal lines outside each box denote the upper and lower whiskers, and dot points denote the outliers. The obvious effect is the reduction in the average variability, with respect to the reliability parameter, i.e., as the number of samples increases (higher reliability values), the results become more stable, which we believe has roots in the central limit theorem.

Considering that most of the published RANSAC-variants used the original RANSAC as main comparison term, and that the proposed method is actually a variant of that method, we decided accordingly and provide the comparison between the performance achieved by RANSAC and the

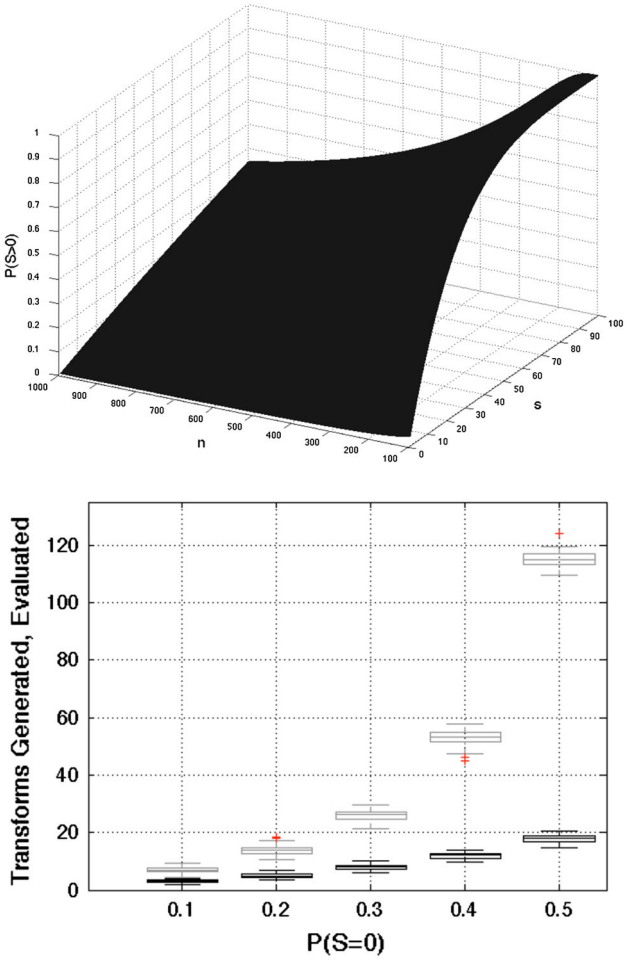

Fig. 7 The top plot gives the probability $P(S > 0)$, i.e., the probability of selecting at least one spurious correspondence when sampling four elements, with respect to n and s. The *bottom plot* gives two boxplot series: the *light boxplots* give the number of hypothesis generated by the proposed method, with respect to the prior probability of drawing a sample composed exclusively of genuine correspondences. The *dark boxplots* give the corresponding sub-sets that were actually taken into account in the evaluation process, which were in the neighborhood of the point of maximum density in eigenspace

Table 1 Parameters of the proposed method and corresponding values used in our experiments

Symbol	Description	Value
α_1	Number of keypoint correspondences	[10, 300]
$\alpha_2 \times \alpha_3$	Dimensions of the correspondences matrix	8×9
$\alpha_4 \times \alpha_4$	Dimensions of the homography matrix	3×3
α_5	Number of image pixels per image	480, 000

given method. Also, by sharing the RANSAC comparison term, it will be straightforward to compare the performance of the proposed method to other RANSAC variants. Figure 10 gives a comparison between the ROC curves obtained by the proposed method (dark lines) and RANSAC (red lines) in our data sets. The vertical lines correspond to the maximum and minimum observed performance at each operating point, when repeating the evaluation process on sub-sets (300 images) randomly selected from the original data set. The dashed black line corresponds to the TPR = FPR curve, i.e., to a totally random pattern recognition system. The sensitivity values observed at the different false match levels were quite close between both algorithms, with a slight decrease in performance of the method described in this paper (less than 1%, on average). This was mainly explained by the reduced number of hypothesis evaluated on each iteration that might neglect some valid transforms sometimes. Also, a slightly higher variability of the results of the proposed method on different sub-sets was observed, which again is justified by the dramatically lower number of transforms evaluated, when compared to the RANSAC strategy. In this case, we observed reductions of more than two orders of magnitude in the number of hypothesis that were considered for the phase of the proposed method that is more concerning in terms of the computational burden (apply homographies to data). This improvement in performance is specially evident in cases where the proportion of spurious correspondences is low.

4 Conclusions and further work

The use of keypoint-based detection and matching techniques has been gaining popularity for a broad range of scenarios. However, the occurrence of spurious correspondences between keypoints is unavoidable and a common source of errors. In this paper, we proposed a stochastic method that focuses on the spatial distribution of keypoint correspondences on non-deformable planar objects to estimate the goodness of correspondences, discriminating between *spurious* and *genuine*. Being based on the RANSAC algorithm, the key novelty is that genuine correspondences tend to agree in the eigenvectors of the homographies that map data, in contrast to spurious correspondences. By finding the most likely *true* homography, the reliability of each keypoint correspondence is found. Our experiments confirmed the effectiveness of the proposed technique, even when the proportion of spurious correspondences is around the classical breakdown value for the proportion of spurious correspondences (0.5). We note that RANSAC is able to handle more than half of spurious correspondences, but in such cases the algorithm starts to converge in a much more slowly way.

For the moment, experiments were constrained to images that contain a single planar surface, which lead to single peaked density estimates in the eigenvectors space. In the future, we expect to extend our work to images with multiple planar surfaces that appear to lead to multiple peak density estimates in the eigenvectors space. This point will be subject of further analysis. Also, another possible extension for this

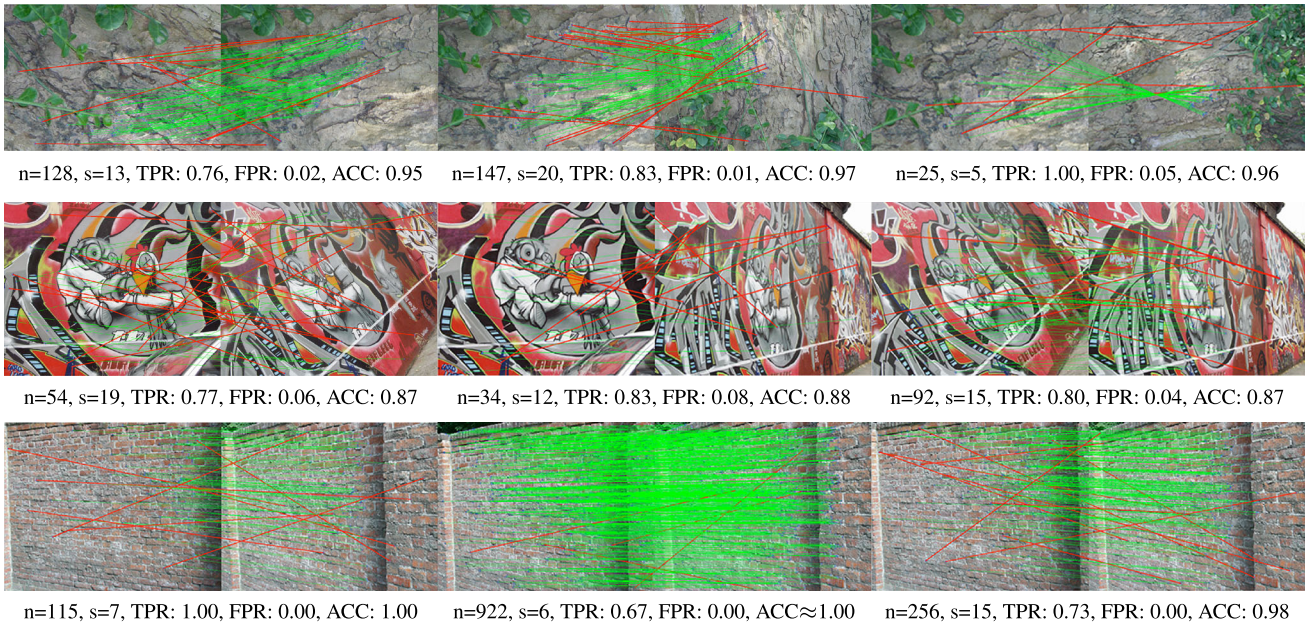

Fig. 8 Examples of the discrimination obtained between the genuine (*thin green lines*) and spurious (*thick red lines*) keypoint correspondences, using the *bark*, *wall* and *graf* sets of the Oxford VGG data

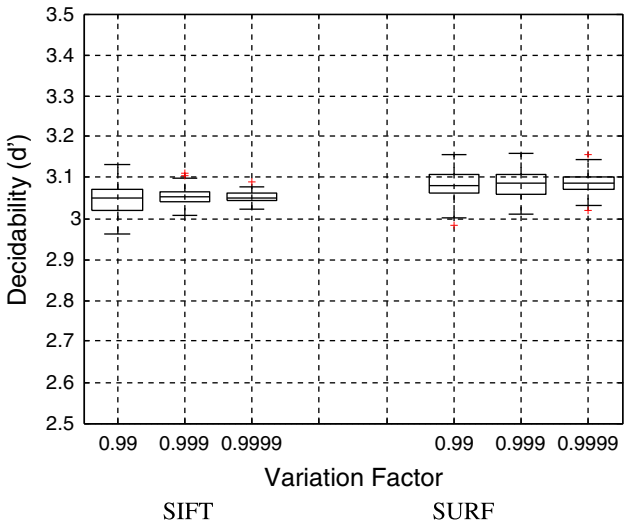

Fig. 9 *Boxplot* of the variations in the decidability score (16) obtained when repeating the stochastic process described in this paper, with respect to reliability parameter

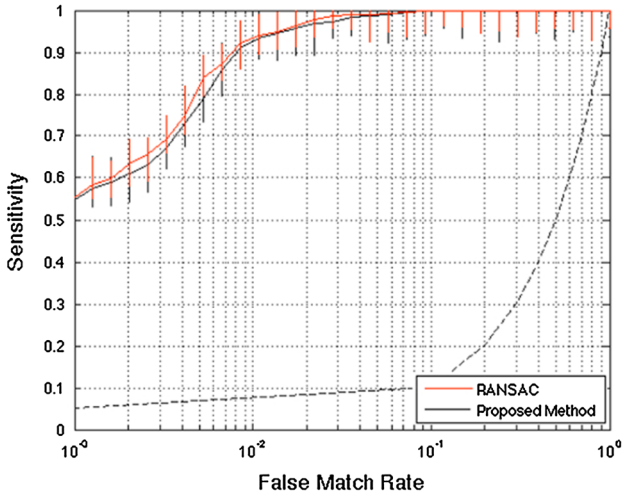

Fig. 10 Comparison between the performance of classical RANSAC discrimination strategy and the proposed method, using SIFT keypoints detection and matching strategies in the full data set

work is the study of *second-order* algorithms to select only a subset of the transforms in the most dense subspace, instead of selecting all of them, as it is suggested in this paper. A possibility might be to use algorithms such as the Mean shift [7], even considering that this phase will essentially speed-up the discrimination process, at expenses of a slight decrease in the accuracy obtained.

Acknowledgments The financial support given by "FCT-Fundao para a Cincia e Tecnologia" and "FEDER" in the scope of the PTDC/EIA/103945/2008 research project "NECOVID: Negative Covert Biometric Recognition" is acknowledged. Also, the support given by IT-Instituto de Telecomunicaes in the scope of the "NOISYRIS" research project is also acknowledged.

References

1. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346–359. (2008)
2. Box, G., Cox, D.: An analysis of transformations. J. Roy. Stat Soc Ser B **26**(2), 211–252 (1964)

3. Calonder, M., Lepetit, V., Ozuysal, M., Trzcinski, T., Strecha, C., Fua, P.: BRIEF: computing a local binary descriptor very fast. IEEE Trans. Pattern Anal. Mach. Intell. (2012). doi:10.1109/TPAMI.2011.222
4. Chin, T.-J., Yu, J., Suter, D.: Accelerated hypothesis generation for multi-structure data via preference analysis. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 625–638 (2012)
5. Choi, S., Kim, T., Yu, W.: Performance evaluation of RANSAC family. In: Proceedings of the British Machine Vision Conference, pp. 1–12 (2009)
6. Chum, O., Matas J.: Matching with PROSAC—progressive sample consensus. In: Proceedings of the 2005 Conference on Computer Vision and Pattern Recognition, (1), pp. 220–226. (2005)
7. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
8. Du, Y., Belcher, C., Zhou, Z.: Scale invariant Gabor descriptor-based noncooperative Iris recognition. EURASIP J. Adv. Signal Process. (2010). doi:10.1155/2010/936512
9. Duy-Nguyen, T., Wei-Chao, C., Gelfand, N., Pulli, K.: SURFTrac: efficient tracking and continuous object recognition using local feature descriptors. In: Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition, (2009). doi:10.1109/CVPR.2009.5206831
10. Fischler, M.A., Bolles R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
11. Golub, G., Van Loan, C.: Matrix Computations, 3rd edn. Johns Hopkins University Press (1996) ISBN: 0-8018-5414-8
12. Goshen, L., Shimshoni, I.: Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. Int. J. Comput. Vis. **80**(2), 275–288 (2008)
13. Hanajik, M., Ravas, R., Smiesko, V.: Interest point detection for vision based mobile robot navigation. In: Proceedings of the IEEE 9^{th} International Symposium on Applied Machine Intelligence and Informatics, pp. 207–211. (2011)
14. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0-521-54051-8, 2003
15. Huang, C., Chen, C., Chung, P.: Contrast context histogram: an efficient discriminating local descriptor for object recognition and image matching. Pattern Recogni. **41**, 3071–3077 (2008)
16. Kenney, C.S., Manjunath, B.S., Zuliani, M., Hewer, G., Van, A.: Novel. A condition number for point matching with application to registration and post-registration error estimation. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1437–1454 (2004)
17. Liu, C., Yuen, J., Torralba, A., Freeman, W.: SIFT flow: dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. **33**(5), 978–994 (2011)
18. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
19. Matej, K., Ales, L., Danijel, S.: Multivariate online Kernel density estimation with Gaussian Kernels. Pattern Recognit. **44**(10–11), 2630–2642 (2011)
20. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
21. Ni, K., Jin, H., Dellaert, F.: GroupSAC: efficient consensus in the presence of groupings. In: Proceedings of the International Conference on Computer Vision, pp. 2193–2200. (2009)
22. Ozuysal, M., Calonder, M., Lepetit, V., Fua, P.: Fast keypoint recognition using random ferns. IEEE Trans. Pattern Anal. Mach. Intell. **32**(3), 448–461 (2010)
23. Nistèr, D.: Preemptive RANSAC for live structure and motion estimation. Mach. Vis. Appl. **16**(5), 321–329 (2005)
24. Quelhas, P., Monay, F., Odobez, J.-M., Gatica-Perez, D., Tuytelaars, T.: A thousand words in a scene. IEEE Trans. Pattern Anal. Mach. Intell. **29**(9), 1575–1589 (2007)
25. Sande, K., Gevers, T., Snoek, C.: Evaluating color descriptors for object and scene recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1582–1596 (2010)
26. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vis. **37**(2), 151–172 (2000)
27. Shi, J., Tomasi, C.: Good features to track. In: IEEE Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 593–600. (1994)
28. Tola, E., Lepetit, V., Fua, P.: DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 815–830 (2010)
29. Tordoff, B., Murray, D.: Guided-MLESAC: faster image transform estimation by using matching priors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1523–1535 (2005)
30. Torr, P.H.S., Zisserman, A.: MLESAC: a new robust estimator with application to estimating image geometry. Comput. Vis. Image Underst. **78**, 138–156 (2000)

Hugo Proença received the B.Sc. degree from the University of Beira Interior, Portugal, in 2001, the M.Sc. degree from the Faculty of Engineering, University of Oporto, in 2004, and the Ph.D. degree from the University of Beira Interior, in 2007. His research interests are focused in the artificial intelligence, pattern recognition and biometrics. Currently, he serves as Assistant Professor in the Department of Computer Science, University of Beira Interior. He is a Senior Member of the IEEE, the area editor (ocular biometrics) of the IEEE Biometrics Compendium Journal and a member of the Editorial Board of the International Journal of Biometrics. Also, he served as Guest Editor of special issues of the Pattern Recognition Letters, Image and Vision Computing and Signal, Image and Video Processing journals.