Local image Features for Shoeprint Image Retrieval

H. Su, D. Crookes, A. Bouridane, and M. Gueham
School of EEE and CS, Queen's University Belfast
Email: {h.su, d.crookes, a.bouridane, mgueham02}@qub.ac.uk

Abstract

This paper deals with the retrieval of scene-of-crime (or scene) shoeprint images from a reference database of shoeprint images by using a new local feature detector and an improved local feature descriptor. Our approach is based on novel modifications and improvements of a few recent techniques in this area: (1) the scale adapted Harris detector, which is an extension to multi-scale domains of the Harris corner detector; (2) automatic scale selection by the characteristic scale of a local structure. (3) SIFT (Scale-Invariant Feature Transform), one of the most widely investigated descriptors in recent years. Like most of other local feature representations, the proposed approach can also be divided into two stages: (i) a set of distinctive local features are selected by first detecting scale adaptive Harris corners where each of them is associated with a scale factor, and then selecting as the final result only those corners whose scale matches the scale of blob-like structures around them. Here, the scale of a blob-like structure is detected by the Laplace-based scale selection. (ii). for each feature detected, an enhanced SIFT descriptor is computed to represent this feature. Our improvements lead two novel methods which we call the Modified Harris-Laplace (MHL) detector, and the enhanced SIFT descriptor.

In this paper, we demonstrate the application of the proposed scheme to the shoeprint image retrieval problem using six sets of synthetic scene images, 50 images for each, and a database of 500 reference shoeprint images. The retrieval performance of the proposed approach is significantly better, in terms of cumulative matching score, than the existing methods used in this application area, such as edge directional histogram, power spectral distribution, and pattern & topological spectra.

Key words: Shoeprint image, Forensic application, Local image features, Content-based image retrieval.

1 Introduction

Shoeprint Classification Shoeprints are often found at crime scenes and provide valuable forensic evidence. It has been estimated that more than 30% of all burglaries provide usable shoeprints that can be recovered from the crime scene [1]. Because of the pattern of repeated offences, rapid classification of such shoeprints would enable investigating officers not only to link different crimes, but to identify potential suspects while the crime is still 'hot'.

The objectives of the forensic investigator are usually: (i) to determine the make and model of a shoe; (ii) to determine if a particular shoeprint was made by a specific shoe; (iii) to match the shoeprint with other shoeprints, possibly from other crime scenes. An image of a shoeprint is first obtained using a technique such as photography, gel or electrostatic lifting, or by making a cast when the impression is in soil, snow or sand. A matching of the shoeprint is then made against specific image databases containing current and previous shoeprint images and/or images of shoeprints found at other crime scenes, leading to a few of candidate images, which are then used in the second stage. This matching process is usually in two stages: an initial classification which results in a hitlist of likely candidates, followed by a visual identification from this reduced set. For this paper, we are concerned with automating the first classification process.

The study of automatic shoeprint classification and retrieval is still relatively new and immature. Some work has recently been reported in [2-4, 28, 29]. Usually, these algorithms first automatically extract a set of feature vectors from a shoeprint image using different kinds of descriptors, such as Fourier descriptors, Fractals, edges, or pattern and topological spectra, and then sort a database of shoeprints in response to a sample shoeprint. However, several practical difficulties exist which hinder the effectiveness of real world shoeprint classification – both manual and automated, such as capture device-dependent noise, distortions, and cutting-out. In this paper, we generate six sets of synthetic scene images to simulate scene images, from a 'clean' dataset with 500 shoeprints by adding Gaussian noise or scene backgrounds, cutting-out, rescaling, rotating, or selecting salient patterns.

Local Feature Descriptors These are computed from distinctive local regions, and do not require segmentation beforehand. They have proved to be very successful in the applications such as image retrieval and matching [5-8], object recognition and classification [9-13], and wide baseline matching [14]. Consequently, many different scale- and affine-invariant local feature detectors, robust local feature descriptors, and their evaluations have been widely investigated in the literature [15-25].

Typically, a local image feature should have four properties: locality, repeatability, distinctiveness, and robustness to different degradations. The above studies suggest that no single local feature representation so far can beat others in terms of all the above four properties, so a good local feature representation should be a trade off of these properties. The work described in this paper firstly detects a set of distinctive local features from an image by combining a scale adaptive Harris corner detector with a Laplace-based scale selection. Then, for each local feature, an enhanced SIFT [18] descriptor is computed to represent the feature. Finally, the matching of descriptors in two images has been conducted by combining the nearest neighbour with the threshold-based screening, i.e. two descriptors are matched only if one is the nearest neighbour to the other one in the feature space, and at the same time the distance between them is less than a threshold. The distance between two images is computed from only the matched pairs.

The main contributions of this paper are: first, a new robust and distinctive local feature detector, Modified Harris-Laplace; second, an enhanced SIFT descriptor has been presented for the local features detected at the first stage; finally, we have investigated the application of these new local image features to the forensic application of automated shoeprint image retrieval, and compare this technique with other three up-to-date methods on the same sets of shoeprint images, and in terms of the same evaluation measure.

This paper is organized as follows: In section 2, we present the new local feature detector based on the adaptive Harris corner detector and the Laplace-based automatic scale selection. Section 3 introduces the enhanced SIFT descriptor. In section 4, we present the image databases used in this work, and also the experiments of shoeprint image retrieval based on local image features. We then summarize the paper in section 5.

2. Modified Harris-Laplace Detector

A local feature here refers to any point and its neighbourhood in an image where the signal changes significantly in terms of two dimensions. Conventional "corners", such as L-corners, T-junctions and Y-junctions satisfy this, but so do the isolated points, the endings of branches and any location with significant 2D texture. Further, all of these local structures have their characteristic size. K. Mikolajczyk et al in [6] have proposed a new Harris-Laplace detector by using the benefit from the Harris corner detector [26] (high accuracy of location) and that from LoG [27] (robust scale selection). But the way they combine the two powerful techniques does not necessarily result in an accurately located and stable scaled local feature detector, since this detector is actually to check if the response of the Harris measure reaches a maximum at the spatial domain and so does the response of the LoG at the same location along the scale direction. In most cases, the unstable component of this detector is the scale selection, since the stability of the scale selection based on LoG is conditioned that this measure should be computed at the centre of a blob structure. In this section, we shall present our solution to combine the two powerful techniques. Considering the Harris-Laplace detector in [6], we name our detector Modified Harris-Laplace.

The scale adaptive Harris detector [6] is based on an extension of the second moment matrix as shown Formula (1), where, σ_i is the integration scale, σ_d is the differentiation scale and f_α is the derivative computed in the direction of α. The strength of a scale adaptive corner can be measured by $\det(A(x,\sigma_i,\sigma_d)) - \kappa \cdot trace^2(A(x,\sigma_i,\sigma_d))$.

$$A(x,\sigma_i,\sigma_d) = \sigma_d^2 \cdot g(\sigma_i) * \begin{bmatrix} f_x^2(x,\sigma_d) & f_x f_y(x,\sigma_d) \\ f_y f_x(x,\sigma_d) & f_y^2(x,\sigma_d) \end{bmatrix} \tag{1}$$

$$LoG(x,\sigma) = \sigma^2 \left| L_{xx}(x,\sigma) + L_{yy}(x,\sigma) \right| \tag{2}$$

An example of the scale-adaptive Harris corner detection on a synthetic image is shown in Fig. 1 (a), where the figure suggests that for a corner, there might be a series of points detected by this approach due to the diffusion and the spatial extent of the corner. Obviously, only a few of them (normally one or two) present the characteristic size of this corner.

The normalized LoG response of an image is defined as Formula. (2), where σ is the scale, and $L_{\alpha\alpha}$ is the second derivative computed in the direction of α. The LoG based automatic scale selection is based on the fact that the characteristic scale of a local structure very often corresponds to the local extremum of the LoG along the scale direction at the centre of a local blob (see the right curve of Fig. 1 (b)). But it must be noted that in most cases, this principle does not work well for the structures like corners. In Fig. 1 (b), the middle curve shows the LoG responses over scales, at the top-left corner of a square. The figure clearly illustrates why the scale selection of the Harris-Laplace detector [6] is unstable. There are too many extrema on the curve of the LoG responses, resulting in redundant and unstable corners.

However, Fig. 1 also suggests that the scale of the corners, located around a blob, might be related to the scale of that blob, which can be selected by referring to the LoG response. In real world, it is reasonable, in most cases, to assume that a corner can be associated with a blob structure around this corner, since any corner is just a part of an object. Based on this assumption, only those candidate corners, whose scale has a predefined relationship with the scale of a blob structure around them, can be selected as a characteristic scale of a corner. Here, two factors should be considered for this strategy: one is the searching region, the other one is the relationship between the scale of a blob structure and that of a corner. Fig. 2 illustrates this strategy, where the long-dashed circle (r) denote a blob structure, and the short-dashed circles

denote the search region with the radius of r_1 and the reference circle with the radius of r_0. The solid circles are the candidates of the same corner located at the up-left of the square, and the bold solid circle is eventually selected as the characteristic scale of this corner. In all of our experiments, Formula (3) is applied to relate the reference scale r_0 and the search region r_1 with the blob scale r. (This formula is derived by assuming a square blob)

$$\sqrt{2} \cdot r_0 = r = \frac{\sqrt{2}}{2} \cdot r_1 \tag{3}$$
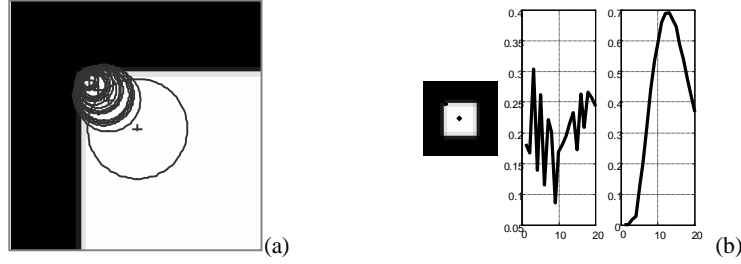
Figure 1 (a), an example of adaptive Harris corner detection, and LoG based scale selection. (b), the responses of the LoG measures at different locations along the scale direction (x-axis) on a synthetic image at the left. The middle curve corresponds to the top-left corner in the left image, and the right curve corresponds to the centre of the square in the same image.
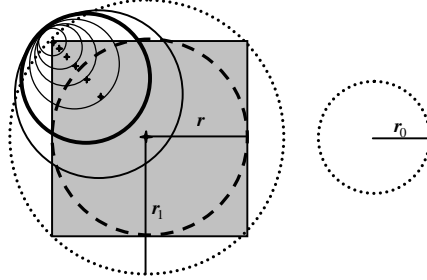
Figure 2 An illustration of automatic scale selection based on the scale of a blob-like structure

3. Local Feature Descriptors

The local photometric descriptors computed in this paper, as mentioned in the first section, are a further extension of the GLOH (Gradient Location and Orientation Histogram) [23, 25], originally from SIFT descriptors. Our method is different from GLOH in that:

(i) First, we apply a circular binary template on each normalized local region to increase the rotation invariance of the descriptor. This is a minor change but has a significant impact on the robustness of the descriptors. Either SIFT or GLOH obtains the rotation invariance by weighting the local region with a Gaussian window. However, it is very often the case that when one chooses a larger sigma of the Gaussian kernel, the descriptors computed on this region are distinctive but rotation sensitive; on the other hand, when one chooses a smaller sigma of the Gaussian kernel, the descriptors are rotation invariant but not distinctive. In most cases, it is hard to choose a proper sigma of the Gaussian window. Therefore we apply a binary template to model the region as a circular one, and meanwhile use a larger sigma of the Gaussian window to keep the distinctiveness of the descriptors.

(ii) Second, we bin the histogram with the orientation range of 180° rather than the original 360° for inverse contrast robustness. In the application of shoeprint image matching, it is very

often that two images contain the same print but with inverse contrast, i.e. in one image the grey level of the print itself is larger than that of the background, whereas in the other image it is otherwise. This difference can be overcome by binning the histogram with the orientation range of 180°, i.e. without considering the polarity.

The construction of our local descriptors is similar to GLOH, i.e. we apply to bin the gradients a log-polar location grid with three bins in the radial direction and four bins in the angular direction (the central grid does not apply angular binning), resulting in a 9 location grid. Considering the orientation range of 180°, 4 bins are applied in gradient orientation. Finally, the descriptors of an image are composed by a $N \times 36$ matrix, where N is the number of the local features detected in the working image.

The similarity between two images depends on the matching strategy of the local features. For the sake of retrieval speed, we apply the nearest neighbour and thresholding jointly to compute the distance between two images, i.e. for each descriptor in one image, the nearest neighbour in another image is found as a potential match, then only those matches whose distance is below a threshold are selected as the final matches. The similarity of two images is computed from the summation of $\exp(-d)$, where d denotes the distances of the final matches.

Of course, there are many other strategies for computing the similarity or matching score between two images. For instance, the nearest neighbour can be applied to obtain the initial matches, and then the RANSAC (Random Sample Consensus) method is used to fit the geometric transformation and reject the mismatches, here the geometric transformation can be a fundamental matrix.

4. Experiments

4.1 Datasets

Our experiments and evaluations for shoeprint image retrieval are conducted on a database of 500 reference images, termed dClean. To simulate the scene images, a few of degraded image sets, have been generated from this base dataset.

(a) dNoisy - Gaussian noise of five different levels is added to each shoeprint in the base dataset. The noise level (σ) varies from 10 to 50 with a step of 10, where the range of the grey levels is from 0 to 255.

(b) dRotate - We have rotated each shoeprint in the base dataset with five random orientations in the range of 0°-90°. The selection of this range, but not 0°-360°, is based on the fact that the algorithms developed in this paper are flip invariant both in horizontal and vertical directions

(c) dRescale - We have rescaled each shoeprint in the base dataset with five random scale ratios in the range of 0.35 to 0.65.

(d) dPartial - Five partial shoeprint images are generated for each shoeprint in the first dataset. The percentage of the partial shoeprint which remains varies from 40% to 95% .

(e) dScene - We simulate the scene images from the base dataset. For each image in the base dataset, we randomly select five scene images from F&F Ltd as the background, and put the "clean" shoeprint on each of the background scenes, resulting in 2500 scene images.

(f) dComplex - This dataset contains 50 shoeprint images with complex backgrounds, geometric transformations, significant "cut-out", or their combinations.

4.2 Other methods

Edge Directional Histogram (EDH) - The authors in [29] claimed that the interior shape information of a shoeprint image was described in terms of its significant edges. Therefore, they applied a histogram of the edge direction as the signature of a shoeprint image. This method first extract the edges using a Canny edge detector, then a total of 72 bins are used to bin the edge directions quantized at 5° intervals. To obtain rotation invariance, they compute the 1-D FFT of the normalized edge direction histogram and take it as the final signature of the shoeprint image.

Power Spectral Distribution (PSD) [4] – this method considers the power spectral distribution as the signature of shoeprint image. To compute the PSD, one needs to firstly down-sample an input image, and secondly takes a 2-D DFT on the down-sampled image, and then the power spectral distribution is computed, finally, a step of masking is taken to obtain the signature. In the similarity computation, the PSD of a query image has to be rotated 30 times, with 1° for each time, and the similarity takes the largest one over the 30 times. In our experiments, the step of rotation is removed because first it is very time consuming to exhaustively try each rotation, and secondly assuming the rotation degree in the range of 30° is not suitable at all for many cases, finally, from the technical point of views any signature can obtain the rotation invariance by this brutal way.

Pattern & Topological Spectra (PTS) [28] - This method starts by (i) de-noising a shoeprint image by a partial differential equation (PDE) approach; (ii) forming a binary shoeprint image by thresholding using Kittler's thresholding approach; (iii) this binary image is then filtered by a series of morphological operations (open) with size-increasing structuring elements. The pattern spectrum and the topological spectrum are constructed by either measuring the area or the Euler number of the filtered images against the size of the structuring elements.

4.3 Performance evaluation

Cumulative Matching Characteristics (CMC) - This measure is suitable for the case where the reference database contains only one relevant record for the query record, and it was claimed in [4] to be able to answer the question "What is the probability of a match if one looks at the first n percent of the database records?" and it can be estimated by computing the proportion of times during trails of the searching when the relevant record appears in the first n percent of the sorted records in the database. The performance of a retrieval system is reflected as a graph which plots the probability of a match (also called cumulative matching score, vertical axis) against the percentage of the records reviewed (horizontal axis).

4.4 Experimental results

The following experiments are to compare the performance, in terms of CMC, of four signatures: i.e. Edge Directional Histogram (EDH), Power Spectral Distribution (PSD), Pattern & Topological Spectra (PTS), and the Local Image Feature (LIF). The experiments are conducted as follows: we first take a shoeprint image from the six degraded datasets as a query image (considering the computation, we randomly choose 50 images from each of degraded datasets (all 50 images from dComplex), therefore, 50 trials for each degraded dataset), and search against the dataset of dClean. For each dataset, we compute the CMC curve, and the results are shown in Fig. 3. Further to the above quantitative comparisons, a few of retrieval examples are shown in Fig. 4, and table 1 also lists the signature sizes of the compared methods. These results suggest that:

6

(i) For the degradations of Gaussian noising, cutting-out, and rescaling, the signatures of PSD and LIF can almost achieve the perfect results, and LIF can also reach the similar results for the degradations of rotation and scene background addition.

(ii) The performance of EDH and PTS is a little bit worse than that of PSD and LIF for the degradations of Gaussian noising, cutting-out, rescaling, and rotation, but both of them still are efficient approaches considering that the cost (signature size) of the two signatures is significantly less than the other two. Within both of them, PTS beats EDH in most cases, while the only exception is on the rescaled database. This matches the results achieved in [28]. Furthermore, it is a surprise that both PTS and EDH beat PSD on the dataset of dScene, even though they both have much less cost than PSD does.

(iii) The signature of LIF works very well for all kinds of degradations. It clearly outperforms other signatures on the dataset with the hardest complex degradations. However, it also costs much more than the signatures of EDH, and PTS.

5. Conclusion

In this paper, we have proposed a new method, called LIF, which combines two novel improvements of existing techniques: Harris-Laplace detector and SIFT descriptor. The new local feature detector, coined Modified Harris-Laplace detector, applies in a new way the scale-adaptive Harris corner detector to determine the candidates of local features, and the Laplace-based automatic scale selection strategy to pick the final local features. We further improve the widely used local feature descriptors – SIFT to be more robust to rotation and inverse contrast. In order to test the performance, we have generated six sets of synthetic scene shoeprint images, and then conducted a few of experiments on shoeprint image retrieval.

The experimental results indicate that, considering the cost of a signature, PTS is a good choice for minor degradations. However, for the real scene images, which are of very bad quality, the proposed signature of LIF based on the Modified Harris-Laplace detector and the enhanced SIFT descriptor is no doubt the best choice for archiving them.

Further issues to be investigated include: (i) further reduce the dimensions of a local feature descriptor; (ii) a fast and accurate matching strategy is necessary for a large shoeprint image database. The current matching based on the nearest neighbour and threholding is fast but not accurate enough, while a more accurate matching strategy based on RANSAC is time-consuming; (iii) there is continuing scope for more advanced local feature detectors and descriptors.

6. Acknowledgements

References

[1] G. Alexandre, "Computerized classification of the shoeprints of burglar's soles", Forensic Science International, Vol. 82, pp. 59-65, 1996.
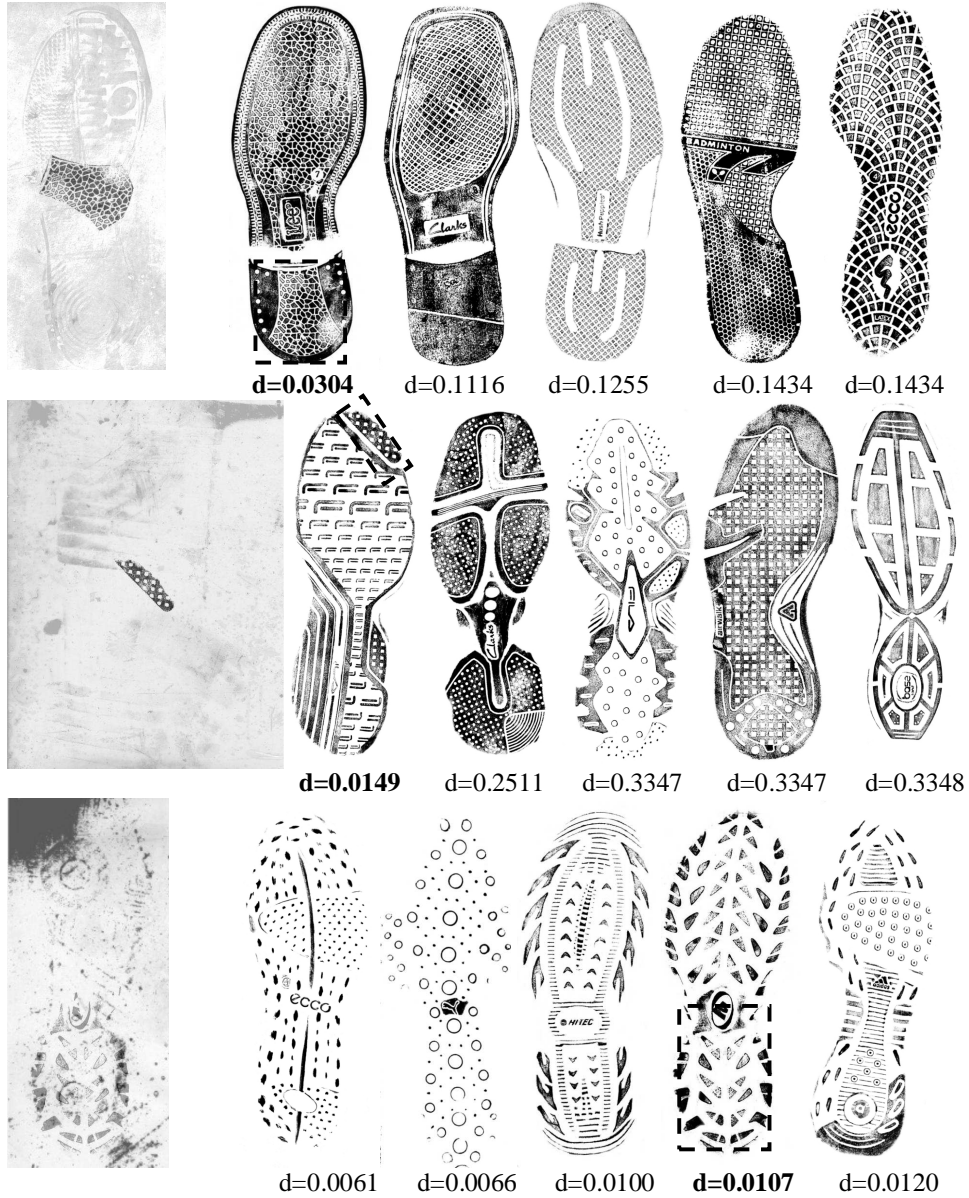
[2] Z. Geradts and J. Keijzer, "The image data REBEZO for shoeprint with developments for automatic classification of shoe outsole designs", Forensic Science International, Vol. 82, pp. 21-31, 1996.

[3] A.G. Alexander, A. Bouridane, D. Crookes, "Automatic classification and recognition of shoeprints," Special issue of the information bulletin for shoeprint/toolmark examiners, Vol. 6. No. 1, pp. 91-104, Mar. 2000.

[4] P. D. Chazal, J. Flynn, and R.B. Reilly, "Automated processing of shoeprint images based on the Fourier transform for use in forensic science," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 3, pp. 341-350, Mar. 2005.

[5] C. Schmid and R. Mohr, "Local greyvalue invariants for image retrieval," IEEE Trans. on PAMI, Vol 19-5, pp. 530-535, 1997.

[6] K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points" ICCV'01, pp. 525—531, 2001.

[7] V. Gouet and N. Boujemaa, "Object-based queries using colour points of interest," IEEE Workshop on Content-Based Access of Image and Video Libraries (CVPR/CBAIVL) Hawai,USA, 2001

[8] L. Ledwich, and S. Williams, "Reduced SIFT Features for Image Retrieval and Indoor Localisation," Australasian Conf. on Robotics and Automation, Canberra, 2004.

[9] D. Lowe, "Object Recognition from Local Scale-Invariant Features," ICCV'99, pp.1150-1157. 1999.

[10] X. Ma and W. Eric L. Grimson, "Edge-Based Rich Representation for Vehicle Classification", ICCV'05, pp. 1550-5499, 2005.

[11] K. Mikolajczyk, B. Leibe, and B. Schiele, "Local Features for Object Class Recognition," ICCV'05, pp.1792-1799, 2005.

[12] B. Platel, E. Balmachnova, L.M.J. Florack, and B.M. ter Harr Romeny, "Top-Points as Interest Points for Image Matching", ECCV'06, Part I, LNCS 3951, pp. 418–429, 2006.

[13] Ran Gal and Daniel Cohen-Or "Salient Geometric Features for Partial Shape Matching and Similarity," ACM Transactions on Graphics, Vol. 25, No.1, pp.130-150, 2006.

[14] T. Tuytelaars and L. Van Gool. "Matching widely separated views based on affine invariant regions," IJCV, Vol. 59, No. 1, pp. 61–85, 2004.

[15] K. Mikolajczyk and C. Schmid, "Scale and Affine invariant interest point detectors," IJCV, Vol. 60, No. 1, pp.63-86, 2004.

[16] T. Kadir, A. Zisserman, and M. Brady, "An affine invariant salient region detector," ECCV'04, pp. 404-416, 2004.

[17] G. Dorko and C Schmid, "Maximally stable local descriptor for scale selection", ECCV'06, pp.504-516, 2006.

[18] D. Lowe, "Distinctive image features from scale-invariant keypoints", IJCV, Vol.60, No. 2, pp. 91-110, 2004.

[19] H. Bay, T. Tuytelaars, L.V. Gool, "SURF: Speed-Up Robust Features", ECCV'06, pp.404-417, 2006.

[20] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," BMVC, Vol. 1, pp. 384--393, 2002.

[21] C. Schmid, R. Mohr, and C. Bauckhage, "Evaluation of interest point detectors", IJCV, Vol. 37, pp. 151 - 172, 2000.

[22] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, L. Gool, "a comparison of affine region detectors", IJCV, Vol. 65, No.1-2, pp.43-72, 2005.

[23] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors," CVPR'03, Vol. 2, pp. 257-264, 2003.

[24] K. Mikolajczyk, C. Schmid, "A performance evaluation of local descriptors," IEEE Trans. on PAMI 27(10):1615-1630. 2005.

[25] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," BMVC, Vol. 1, pp. 384--393, 2002.

[26] C. Harris and M. Stephens, "A combined corner and edge detector," Alvey Vision Conference, pp. 147–151. 1988.

[27] T. Lindeberg, "Feature Detection with Automatic Scale Selection", IJCV, Vol. 30, No. 2, 1998.

[28] H. Su, D. Crookes, and A. Bouridane, "Shoeprint image retrieval by topology and patterns", Queen's University Belfast, Tech. Rep. May 2006

[29] L. Zhang, and N.M. Allinson, "Automatic Shoeprint Retrieval System for use in Forensic Investigations," 5th Annual UK Workshop on Comp. Intelligence, 2005
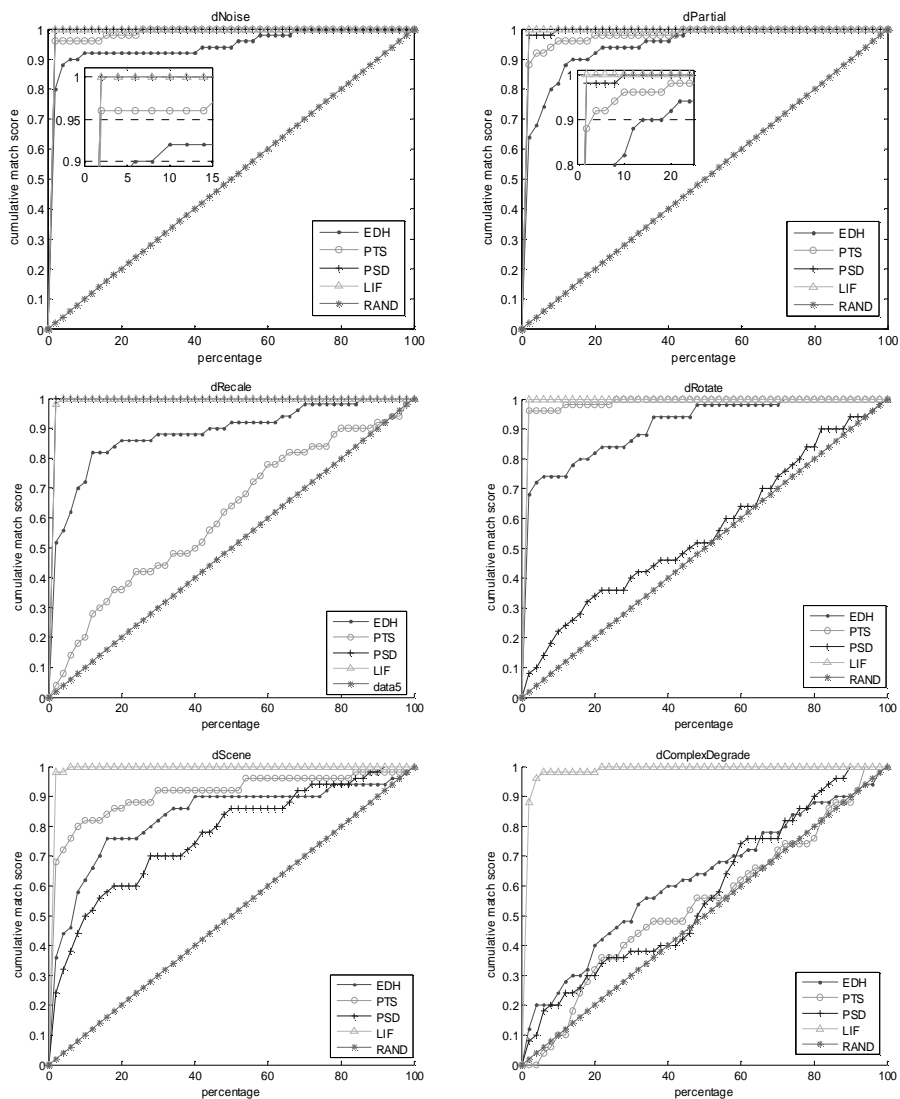
Figure 3 Performance evaluations of four signatures (EDH, PTS, PSD, and LIF) in terms of cumulative matching score on six degraded image dataset. RAND here is the worst case, i.e. the order of the images in the dataset is randomly allocated.

d=0.0304	d=0.1116	d=0.1255	d=0.1434	d=0.1434

d=0.0149	d=0.2511	d=0.3347	d=0.3347	d=0.3348

d=0.0061	d=0.0066	d=0.0100	**d=0.0107**	d=0.0120

Figure 4 Examples of shoeprint image retrieval. In each row, the most left image is a noisy query shoeprint from dComplex, and the rest of the row is the top ranked shoeprint images in dClean. The distance is shown under each retrieved image, and the red squares denote the corresponding patterns contained in the query images.

Methods	EDH	PTS	PSD	LIF
Signature size	72	120	24,099	19,131

Table 1 the signature size of four techniques.