6

# Interest-Point based Face Recognition System

Cesar Fernandez and Maria Asuncion Vicente
*Miguel Hernandez University*
*Spain*

## 1. Introduction

Among all applications of face recognition systems, surveillance is one of the most challenging ones. In such an application, the goal is to detect known criminals in crowded environments, like airports or train stations. Some attempts have been made, like those of Tokio (Engadget, 2006) or Mainz (Deutsche Welle, 2006), with limited success.

The first task to be carried out in an automatic surveillance system involves the detection of all the faces in the images taken by the video cameras. Current face detection algorithms are highly reliable and thus, they will not be the focus of our work. Some of the best performing examples are the Viola-Jones algorithm (Viola & Jones, 2004) or the Schneiderman-Kanade algorithm (Schneiderman & Kanade, 2000).

The second task to be carried out involves the comparison of all detected faces among the database of known criminals. The ideal behaviour of an automatic system performing this task would be to get a 100% correct identification rate, but this behaviour is far from the capabilities of current face recognition algorithms. Assuming that there will be false identifications, supervised surveillance systems seem to be the most realistic option: the automatic system issues an alarm whenever it detects a possible match with a criminal, and a human decides whether it is a false alarm or not. Figure 1 shows an example.

However, even in a supervised scenario the requirements for the face recognition algorithm are extremely high: the false alarm rate must be low enough as to allow the human operator to cope with it; and the percentage of undetected criminals must be kept to a minimum in order to ensure security. Fulfilling both requirements at the same time is the main challenge, as a reduction in false alarm rate usually implies an increase of the percentage of undetected criminals.

We propose a novel face recognition system based in the use of interest point detectors and local descriptors. In order to check the performances of our system, and particularly its performances in a surveillance application, we present experimental results in terms of Receiver Operating Characteristic curves or ROC curves. From the experimental results, it becomes clear that our system outperforms classical appearance based approaches.

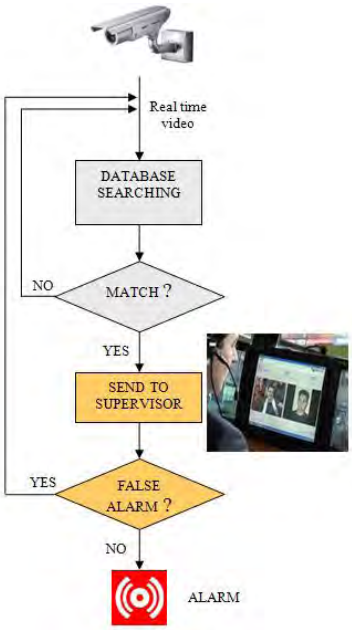


Fig. 1. Example of a supervised surveillance system.

## 2. Previous approaches

Classical face recognition systems are based on global appearance-based methods: PCA or Principal Component Analysis has been used by (Kirby & Sirovich, 1990) and (Turk & Pentland, 1991); ICA, or Independent Component Analysis has been used by (Bartlett et al., 2002), (Draper et al., 2003) and (Liu, 2004). Finally, LDA or Linear Discriminant Analysis has been used by (Belhumeur et al., 2002).

As an alternative to appearance-based methods, local description methods are currently an area of active research in the face recognition field. From Lowe's work on object recognition using SIFT (Scale Invariant Feature Transform) descriptors (Lowe, 2004), multiple authors have applied such descriptors in other fields, like robot navigation (Se et al., 2001), scene classification (Pham et al., 2007), and also face recognition.

Some of the main contributions using SIFT descriptors for face recognition will be briefly described: Lowe (Lowe, 2000) presents a similar scheme to that of object recognition, but does not address the problem of face authentication. Sivic (Sivic et al., 2005) combines PCA and SIFT: PCA is used to locate eyes, nose and mouth; while SIFT descriptors are used to describe fixed-sized areas around such points. Finally, Bicego (Bicego et al., 2006) measures the distance between two faces as the distance of the best matching pair of descriptors, in some cases using previous knowledge about the location of eyes and mouth.

The goal of our work is to propose a new distance measure in order to exploit the potential of SIFT descriptors in the face recognition field.

## 3. Interest point detection

Interest point detectors try to select the most descriptive areas of a given image. Ideally, given multiple images of the same object or person, under different lighting, scale, orientation, view angle, etc., a perfect algorithm would find exactly the same interest points across all images.

In the field of face recognition, although invariance to orientation and view angle are necessary, images that are useful for face recognition always present the user from the similar angles (usually, facing the camera) and orientations (standing up). Possible view angles and orientations are expected to be within a 30 degree range, approximately. Interest point detectors that allow much higher ranges of variation are not necessary and more simple, faster detectors would be preferred instead.

In that sense, affine invariant detectors, like those detailed in (Alvarez & Morales, 1997), (Baumberg, 2000) or (Mikolajczyk & Schmid, 2004) are not considered for our work. We have made experiments with two more simple detectors: Harris-Laplace (Mikolajczyk & Schmid, 2004) and Difference of Gaussian (Lowe, 2004).

The Harris-Laplace detector is a scale-invariant version of the well-known Harris corner detector (Harris & Stephens, 1988) and looks for corners or junctions in the images. On the other side, the Difference of Gaussian detector (DoG) is an approximation to the Laplacian of Gaussian operator, and looks for blob-like areas in images. Both detectors have been widely used in the object recognition field and they are highly reliable. In figure 2 we show the interest points found by each of these detectors over the same image (the diameter of the circle is represents the scale of the detected interest area).

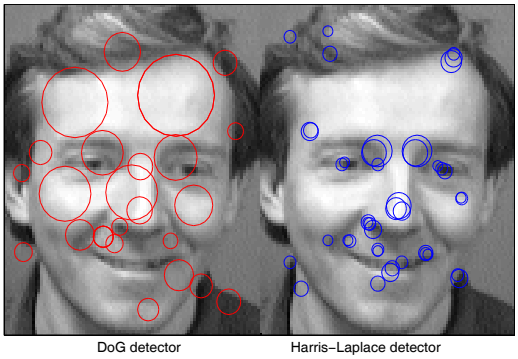


Fig. 2. Output of Harris-Laplace and DoG interest point detectors.

It becomes clear that each detector looks for specific image areas, and that, depending on the particular application, one of them should be preferred. In the case of face recognition, both

sets of interest points seem to be relevant for describing faces, so our option has been to keep all interest points found by both detectors. The goal is to obtain as much information as possible from each image.

## 4. Interest point description

Once interest points are detected, their surrounding area must be encoded or described by a distinctive feature. Ideally, features should be invariant to lighting, scale, orientation, view angle, etc. At the same time, those features should be unique, in the sense that a different area of the object (or face), a different object, or a different person would be distinguishable.

In (Mikolajczyk & Schmid, 2005) a detailed comparison of local descriptors is carried out. The conclusion is that SIFT (Lowe, 2004) and other SIFT-like descriptors, like PCA-SIFT (Ke & Sukthankar, 2004) or GLOH (Mikolajczyk & Schmid, 2005) give the best results throughout all tests. We will briefly describe some of these descriptors.

Basically, in SIFT descriptors, the neighbourhood of the interest point, scaled accordingly to the detector information, is described as a set of orientation histograms computed from the gradient image. SIFT descriptors are invariant to scale, rotation, lighting and viewpoint change (in a narrow range). The most common implementation uses 16 histograms of 8 bins (8 orientations), which gives a 128 dimensional descriptor.

PCA-SIFT descriptor is also based on the gradient image, the main difference with SIFT being the further compression using PCA. The uncompressed dimension of the descriptor is 3042 (39x39), which is reduced to 36 after applying PCA. The authors claim improved accuracy and faster matching, but these performance improvements are not consistent throughout all tests, as it is shown in (Mikolajczyk & Schmid, 2005).

GLOH stands for Gradient Location-Orientation Histogram. It is also a SIFT-based descriptor, with modified location grids (both polar and Cartesian location grids are considered) and a further PCA compression of the information, which keeps the 128 largest eigenvectors (the dimension of the uncompressed descriptor is 272). GLOH outperforms SIFT in certain situations, with structured scenes and high viewpoint changes. However, such situations are not common in a face recognition scenario.

Recently, the SURF or Speeded Up Robust Features descriptor (Bay et al., 2006) has appeared as an alternative to SIFT. Its main advantage is its fastest computation, while keeping a high descriptive power. It is partially inspired by SIFT, but instead of using the gradient image, it computes first order Haar wavelet responses. Additionally, the use of integral images is the key factor for fast computation. So far, we have not performed tests with the SURF descriptor, so we cannot affirm its validity for face recognition applications.

Finally, LESH or Local Energy based Shape Histogram descriptor (Sarfraz & Hellwich, 2008), has been specifically designed for face recognition applications. Its goal is to encode the underlying shape present in the image. Basically, the descriptor is a concatenation of histograms obtained by accumulating local energy along several filter orientations.

However, it is focused in pose estimation, so it addresses a different problem to that of our work.

In conclusion, we decided to describe each face image with SIFT descriptors computed at all the interest points found by the Harris-Laplace and the DoG detectors.

## 5. Similarity between two face images

Once we have represented all face images as a set of interest points and their corresponding descriptions, the next step to be carried out is the definition of a similarity measure between two face images, in order to be able to decide whether such images correspond to the same person or not.

The simplest approach is to obtain the best possible correspondence between the interest points of both images (according to the values of their SIFT descriptors) and to compute Euclidean distances between each pair of corresponding points. However, according to Lowe's work (Lowe, 2004), SIFT descriptors must be used in a slightly different way: in order to decide whether two points in two different images correspond or not, the absolute value of the Euclidean distance is not reliable; what should be used instead is the ratio between the best match and the second best match. Briefly, for each point of the first image, the best and second best matching points of the second image must be found: if the first match is much better than the second one (as measured by the ratio between SIFT differences) the points are likely to correspond. Eq. 1 shows how to apply such condition, where points ***B*** and ***C*** in $image_2$ are the best and second best matches, respectively, for point ***A*** in $image_1$.

$$\frac{\left|SIFT^{A}_{image_1} - SIFT^{B}_{image_2}\right|}{\left|SIFT^{A}_{image_1} - SIFT^{C}_{image_2}\right|} < Threshold \qquad A, image_1 \text{ corresponds to } B, image_2 \qquad (1)$$

We have used such approach in order to compute the number of corresponding points between two images, such a number being our first measure of similarity between the images. In our notation, the number of matching points between images ***A*** and ***B***, according to the descriptor values is $\boldsymbol{MD_{AB}}$.

Even though we compute similarity according to Lowe's recommendations, the number of correct matches is not completely reliable as a measure of similarity. We have added two extra measures in order to increase the robustness of our system.

The first extra measure is obtained by computing the number of corresponding points that are coherent in terms of scale and orientation: every detected point output by the Harris-Laplace of DoG detectors has an associated scale and orientation. Scale and orientation may be different between images, even if they belong to the same person, but such difference must be coherent across all matching points. Our second measure of similarity is the number of matching points coherent in terms of scale and orientation (a simple Hough

transform is used to obtain the maximum number of points fulfilling this condition). We will refer to this extra measure as $\boldsymbol{MSO_{AB}}$.

The second extra measure is obtained by imposing an additional restriction: the coherence in terms of relative location of corresponding points. Theoretically, the relative location of all matching points must be similar between two images, even if there are scale, rotation and viewpoint changes between them. We will consider a general affine transformation between images for the sake of simplicity (since faces are not planar, high viewpoint changes cannot be represented by affine transformations). The number of points coherent in the parameters of the transformation will be our third measure of similarity. We will use $\boldsymbol{MRL_{AB}}$ to refer to this measure.

Obviously, whenever an additional restriction is imposed, the robustness of the measure is increased, so the second extra measure is the most robust one, followed by the first extra measure and by the original one. In order to compare whether a certain image ***A*** is more similar to image ***B*** or to image ***C*** (i.e., we are trying to classify image ***A*** as belonging to subject ***B*** or subject ***C***), the decision tree of Fig. 3 should be used:

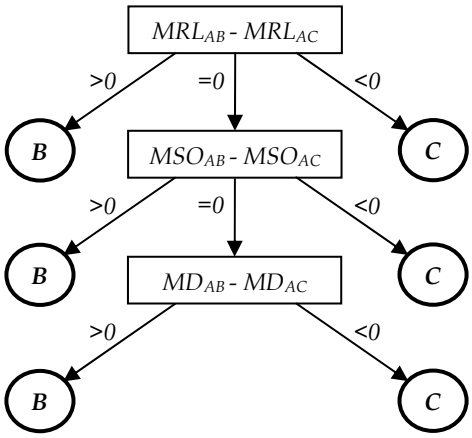


Fig. 3. Decision tree for the classification of image A as belonging to subjects B or C.

Even though a decision tree representation is valid for a classification problem, it cannot be used in an authentication application, where a threshold must be fixed. In order to cope also with such applications, we propose a simple distance measure ***M*** (see eq. 2) that combines ***MRL***, ***MSO*** and ***MD***, giving ***MRL*** a weight one order of magnitude above ***MSO*** and two orders of magnitude above ***MD***.

$$M_{AB} = MD_{AB} + 10MSO_{AB} + 100MRL_{AB} \quad (2)$$

In our experiments, such simple distance measure has shown to give the same results as the decision tree of Fig. 3.

## 6. Experimental results

### 6.1 Databases and baseline used for the evaluation

We have selected two different databases for the evaluation of our face recognition algorithm. The first one is the well-known AT&T database (Samaria, 1994)(AT&T, 2002); and the second one the LFW or Labelled Faces in the Wild database (Huang et al., 2007)(University of Massachusetts, 2007).

The AT&T database contains 40 subjects, each of one described by 10 frontal face images. All images were taken under controlled conditions of lighting, distance to the camera, etc. The main differences between shots are facial expression, and slight orientation and viewpoint changes.

The LFW database contains 5749 subjects, described by a number of images that ranges from 1 to 530. All images have been obtained from the World Wide Web, manually labelled and cropped using the Viola-Jones face detector. Variability between images of the same subject is much higher than that of the AT&T database, thus making LFW more challenging for a face recognition application. For our tests, we have selected a subset containing the 158 subjects described by at least 10 images, and we have kept only the first 10 images of each subject.

As the baseline for our evaluation, we have selected the classic PCA approach to face recognition. We have decided to use PCA because other similar approaches like ICA or LDA have not proved to perform better. In particular in one of our previous papers (Vicente et al., 2007) we showed the equivalence of PCA and ICA under restrictions such as the use of rotational invariant classifiers.

### 6.2 Results

As the goal of our paper is to evaluate our face recognition method for surveillance applications, we have decided to use ROC curves for showing our experimental results. The main idea is to express the relationship between false alarm rates and percentage of undetected criminals. As both databases (AT&T and the LFW subset we are using) share the same structure of 10 images per subject, in both cases we used 4 subjects for training and the remaining 6 subjects for testing. Every test image was compared to all training images of all subjects, the global distance to a subject being computed as the minimum across the 4 training images of such subject (we performed some tests using the mean distance for all training images of the subject, but the results were worse).

First, we performed some experiments in order to adjust our algorithm for the best overall results. The main parameter to tune was the threshold for accepting or rejecting matches between interest points of two different images (see Eq. 1). We carried out tests with both databases (AT&T and LFW) and with thresholds ranging from 0.60 (the most restrictive) to 1.00 (the less restrictive, all matches are accepted).

Figure 4 shows the results obtained with the AT&T database. The left plot shows the full ROC curve, where the different experiments are almost indistinguishable. All of them show

close to ideal behaviours, as it was expected for such database, were all images were taken under controlled conditions. In order to show the differences between the experiments, the right plot shows a scaled detail of the upper left area of the ROC curve. However, all values of the threshold seem to perform similarly and no conclusions can be drawn.

Figure 5 shows the results obtained for the LFW database. In this case, we performed two different experiments. For the first one (left plot), we used the first four images of each subject as training images, keeping the original image order of the database. The results show clearly that the LFW database is much more challenging for a face recognition algorithm, with ROC curves far from the ideal ones. The main reason is the high variability between images of the same subject. For the second experiment (right plot) we decided to rearrange the database, so that the best four images of each subject were selected as training data. Such rearrangement makes the experiment more realistic, since in a real surveillance application training images (taken under controlled conditions) usually have higher quality than test images (taken under uncontrolled conditions, in real time). Anyway, the results are similar in both experiments.

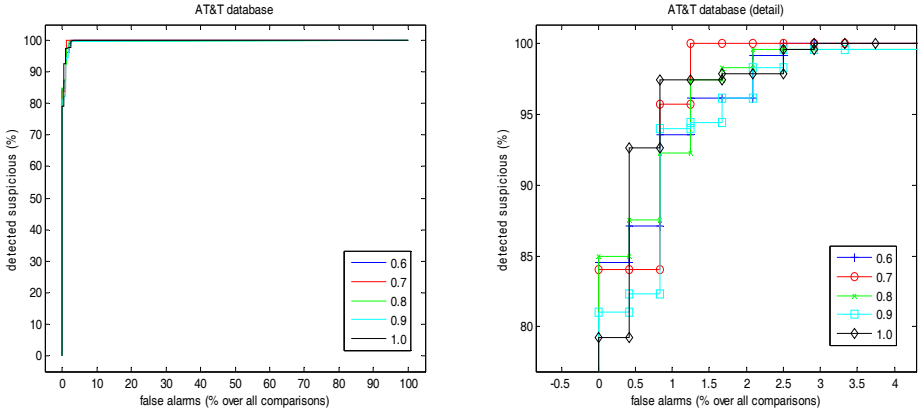


Fig. 4. AT&T database: experiments with different thresholds

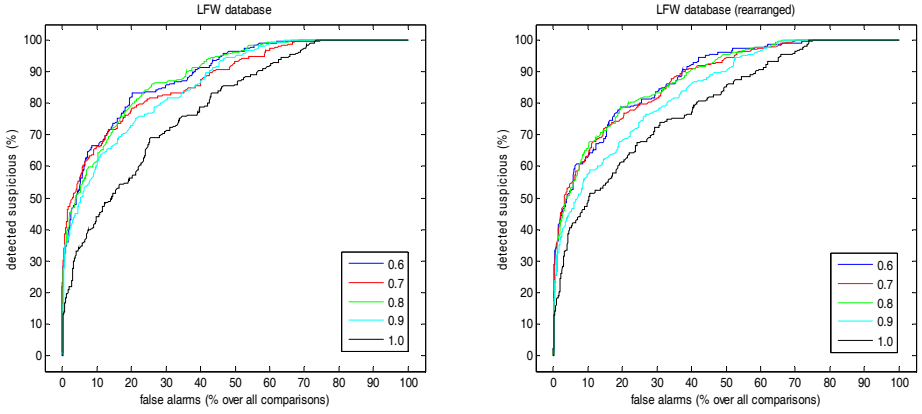


Fig. 5. LFW database: experiments with different thresholds

Concerning the threshold values, both plots of figure 5 show similar behaviours: the results improve as the threshold is reduced, up to a certain point where differences are small (in the range from 0.8 to 0.6). A threshold value of 0.6 seems to perform slightly better than the other settings, so we kept this value for further experiments.

Once the threshold was fixed, we performed several comparisons between our algorithm and the PCA baseline, working with the same databases. Figure 6 shows the results obtained for the AT&T database (left plot) and the LFW database (right plot). Our method clearly outperforms PCA throughout the ROC curve for both databases.

The left plot of figure 7 shows the comparison between PCA and our method for the rearranged version of LFW database (the 4 best images are used for training). There is a slight increase in the performances of both PCA and SIFT, but our method is still clearly superior. Finally, the right plot of figure 7 shows a further experiment: we sorted the 6 test images of LFW for each subject, so that the first 3 images were the best, easier to classify and the last 3 images were the worst, more difficult to classify, according to our opinion. The goal was to check to what extent the performances of both methods were affected by the (subjective) quality of the images: although there are not big differences, it seems that our method is more robust than the PCA approach.

Concerning the feasibility of the proposed approach for a surveillance application, our experimental results show the importance of image quality. The ROC curves obtained for the AT&T database (figure 6, left plot) are close to ideal: at a false alarm rate of 1%, it is expected that 94% of the criminals would be correctly identified (88% at a 0.5% false alarm rate). Such performances would allow us to implement the system in a real scenario. However, the ROC curves obtained for the LFW database, even if the best images are selected for training, are far from ideal: at a false alarm rate of 1%, it is expected that only 35% of the criminals would be correctly identified (32% at a false alarm rate of 0.5%). Such a system would be of little help as a surveillance tool.

As image quality is a key factor for the feasibility of the system, our recommendation is to study properly the location of the video cameras. In our opinion, if video cameras are located in relatively controlled places like walkthrough detectors, the image quality may be enough as for a successful implementation of a supervised surveillance system.

## 7. Conclusion

Automatic or supervised surveillance applications impose strict requirements in face recognition algorithms, in terms of false alarm rate and percentage of undetected criminals. We present a novel method, based on interest point detectors (namely, Harris-Laplace and DoG) and SIFT descriptors.

Our measure of similarity between images is based on computing the number of corresponding points that, apart from having similar values for their SIFT descriptors, fulfil scale, orientation and relative location coherence. Images with a higher number of

corresponding points are likely to belong to the same subject. Such a simple similarity measure has proven to perform consistently in our tests.

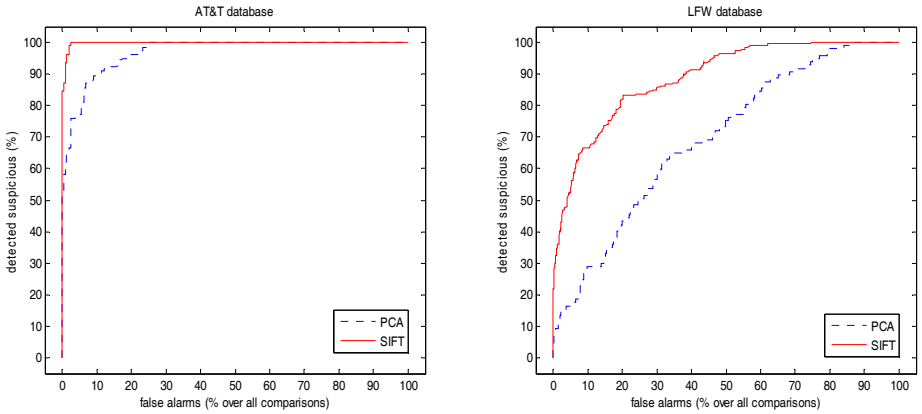


Fig. 6. AT&T and LFW databases: comparison with PCA baseline

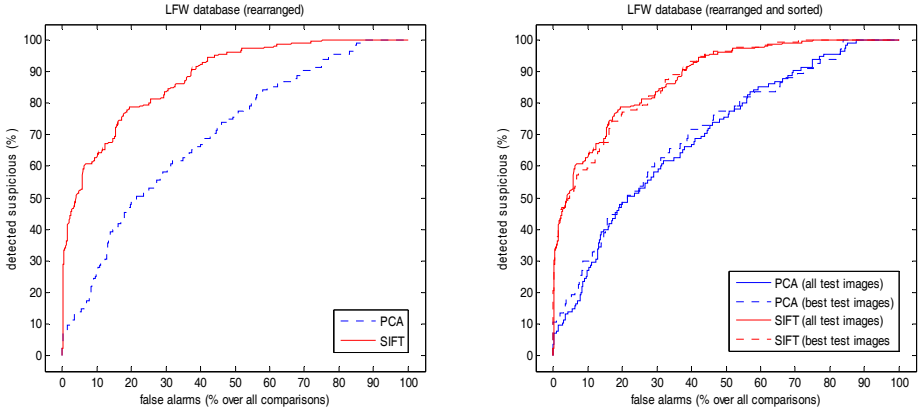


Fig. 7. LFW database reordered and sorted: comparison with PCA baseline

The results in terms of ROC curves show that our approach clearly outperforms the PCA baseline in all conditions. We have performed tests with two different databases: AT&T (not very demanding for a face recognition algorithm) and LFW (extremely demanding); and in both cases our algorithm gave much higher recognition rates than PCA.

Concerning the feasibility of a supervised surveillance system based on our face recognition algorithm, the experimental results show that the quality of the images should be comparable to that of the AT&T database. For lower quality images like those of the LFW database, high recognition rates cannot be expected.

Future work to be carried out includes the comparison of our proposal against other approaches like AAM (active appearance models) and the use of a different interest point descriptor (namely, the SURF descriptor). Another important topic for future research is an evaluation of the possible placements for surveillance cameras; such a research could give us realistic information about the feasibility of a supervised surveillance system.

## 8. References


Alvarez, L. & Morales, F. (1997). Affine morphological multiscale analysis of corners and multiple junctions. *International Journal of Computer Vision,* 25, 2, (November 1997) 95-107, ISSN 0920-5691.

AT&T (2002). The Database of Faces (formerly "The ORL Database of Faces"). *www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html*

Bartlett, M.S.; Movellan, J.R. & Sejnowski, J. (2002). Face recognition by independent component analysis. *IEEE. Trans. Neural Networks,* 13, 6, (November 2002) 1450-1464, ISSN 1045-9227.

Baumberg, A. (2000). Reliable feature matching across widely separated views, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, pp. 774-781, ISBN 0-7695-0662-3, Hilton Head, SC. USA, June 2000, IEEE Computer Society.

Bay, H.; Tuytelaars, T.; Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Proceedings of the 9th International Conference on Computer Vision*, pp. 404-417, ISBN 3-540-33832-2, Graz, Austria, May 2006, Springer-Verlag.

Belhumeur, P.N.; Hespanha, J.P. & Kriegman, D.J. (2002). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Analysis and Machine Intelligence,* 19, 7, (August 2002) 711-720, ISSN 0162-8828.

Bicego, M.; Lagorio, A.; Grosso, E. & Tistarelli, M. (2006). On the use of SIFT features for face authentication, *Proceedings of Conf. on Computer Vision and Pattern Recognition Workshop*, pp. 35, ISBN 0-7695-2646-2, New York, NY, USA, June 2006, IEEE.

Deutsche Welle (2006). German Anti-Terrorist Technology Scans Faces in Train Stations. *www.dw-world.de/dw/article/0,2144,2222284,00.html*.

Draper, B.; Baek, K.; Bartlett, M.S. & Beveridge, R. (2003). Recognizing faces with PCA and ICA. *Computer Vision and Image Understanding,* 91, 1, (July 2003) 115-137, ISSN 1077-3142.

Engadget (2006). Tokyo train station gets facial scan payment systems. *www.engadget.com/2006/04/27/tokyo-train-station-gets-facial-scan-payment-systems/*

Harris, C & Stephens, M. (1988). A combined corner and edge detector, *Proceedings of the 4th Alvey Vision Conference*, pp. 147-151, Manchester, UK, September 1988, University of Sheffield.

Huang, B.; Ramesh, M.; Berg, T. & Learned-Miller, E. (2007). Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, *Technical Report 07-49, University of Massachusetts*, Amherst, October 2007.

Ke, Y. & Sukthankar, R. (2004). PCA-SIFT: A more distinctive representation for local image descriptors, *Proceedings of the Conference on Computer Vision and Pattern Recognition*, pp. 511-517, ISBN 0-7695-2158-4, Washington, USA, June 2004, IEEE Computer Society.

Kirby, M. & Sirovich, L. (1990). Application of the Karhunen-Loeve procedure for the characterization of human faces. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 12, 1, (January 1990) 103-108, ISSN 0162-8828.

Liu, C. (2004). Face Enhanced independent component analysis and its application to content based face image retrieval. *IEEE. Trans. Systems, Man, and Cybernetics, Part B: Cybernetics,* 34, 2, (April 2004) 1117-1127, ISSN 1083-4419.

Lowe, D.G. (2000). Towards a computational model for object recognitionin IT cortex. *Lecture Notes in Computer Science,* 1811, (May 2000) 20-31, ISSN 0302-9743.

Lowe, D.G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision,* 60, 2, (November 2004) 91-110, ISSN 0920-5691.

Mikolajczyk, K. & Schmid, C. (2004). Scale and affine invariant interest point detectors. *International Journal of Computer Vision,* 60, 1, (October 2004) 63-86, ISSN 0920-5691.

Mikolajczyk, K. & Schmid, C. (2005). A Performance Evaluation of Local Descriptors. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 27, 10, (October 2005) 1615-1630, ISSN 0162-8828.

Pham, T.T.; Maillot, N.E.; Lim, J.H. & Chevallet, J.P. (2007). Latent semantic fusion model for image retrieval and annotation, *Proceedings of 16th ACM Conf. on Information and Knowledge Management*, pp. 439-444, ISBN 978-1-59593-803-9, Lisbon, Portugal, November 2007, Association for Computing Machinery, Inc. (ACM).

Samaria, F. & Harter, A. (1994). Parameterisation of a Stochastic Model for Human Face Identification, *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, pp. 235-242, ISBN 978-989-8111-21-0, Sarasola, FL, USA, December 1994.

Sarfraz, M.S. & Hellwich, O. (2008). Head pose estimation in face recognition across pose scenarios, *Proceedings of the Third International Conference on Computer Vision Theory and Applications*, pp. 235-242, ISBN 978-989-8111-21-0, Madeira, Portugal, January 2008, INSTICC - Institute for Systems and Technologies of Information, Control and Communication.

Schneiderman, H. & Kanade, T. (2000). A Statistical Method for 3D Object Detection Applied to Faces and Cars, *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, pp. 1746-1759, ISBN 0-7695-0662-3, Hilton Head, SC. USA, June 2000, IEEE Computer Society.

Se, S.; Lowe, D.G & Little, J. (2001). Vision-based mobile robot localization and mapping using scale-invariant features, *Proceedings of IEEE Int. Conf. on Robotics and Automation*, pp. 2051-2058, ISBN 0-7803-6578-X, Seoul, Korea, May 2001, IEEE.

Sivic, J.; Everingham, M. & Zisserman, A. (2005). Person spotting: Video shot retrieval for face sets. *Lecture Notes in Computer Science,* 3568, (July 2005) 226-236, ISSN 0302-9743.

Turk, M. & Pentland, A. (1991). Eigenfaces for recognition. *J.Cognitive Neuroscience,* 3, 1, (Winter 1991) 71-86, ISSN 0898-929X.

University of Massachusetts (2007). Labeled Faces in the Wild. *vis-www.cs.umass.edu/lfw/*

Vicente, M.A.; Hoyer, P.O. & Hyvarinen, A. (2007). Equivalence of Some Common Linear Feature Extraction Techniques for Appearance-Based Object Recognition Tasks. *IEEE Trans. Pattern Analysis and Machine Intelligence,* 29, 5, (May 2007) 896-900, ISSN 0162-8828.

Viola, P. & Jones, M.J. (2004). Robust Real-Tine Face Detection. *International Journal of Computer Vision,* 57, 2, (May 2004) 137-154, ISSN 0920-5691.