A Fast Scheme for Detecting, Localizing and Grouping Similarly Looking Faces in Random Images

Andrzej SLUZEK
Dept of Electrical and Computer Engineering
Khalifa University
Abu Dhabi, UAE
e-mail: andrzej.sluzek {at} kustar.ac.ae

A scheme is presented for a preliminary detection of similar faces (or face fragments) in unknown images. Another objective of the scheme is to identify groups of images prospectively showing the same person. The scheme employs a standard technique of affine-invariant keypoint detection and matching. However, a novel keypoint descriptor is used which not only characterizes the keypoint itself, but additionally incorporates visual and geometric characteristics of keypoint neighborhoods. Therefore, a straightforward match between two keypoints usually indicates similar areas of significant sizes in compared images. The descriptor is quantized into small vocabularies independently representing photometry and geometry of keypoints. Therefore, a significant amount of visual and geometric distortions can be absorbed and *recall* of face retrieval is reasonably high. However, to keep *precision* at high level as well, we overlap results obtained by using two different detectors, i.e. Harris-Affine and Hessian-Affine. Altogether, the scheme's performances are satisfactorily high (as shown on the test datasets) while its complexity is very low (allowing scalability to large datasets).

Keywords-face identification, face similarity, partial near-duplicates, keypoint matching, keypoint descriptor

I. INTRODUCTION

Preliminary detection of similarly looking faces from images of unpredictable contents can prospectively become a useful tool in security and/or surveillance systems. In such tasks, the objective is usually not to verify the identity of individuals (by comparing a high-quality face photo captured in pre-arranged setups, which is often impossible or inconvenient, against a similarly acquired template). Instead, we attempt to detect similar faces in unspecified collections of provided images. The images may contain faces, but the image quality, the background contents, the view of faces (e.g. partial occlusions) are unpredictable. In particular, there are no template images and the identity of subjects is generally unknown.

Then, the objective is to identify within the provided images faces which look similar (at least partially) so that any further verification (involving more advanced tools or human operators) can be performed on much smaller datasets. Additionally, the images can be grouped into clusters presumably showing the same person.

Such a system should use algorithms of very low complexity (otherwise scalability to very large databases is problematic) even if it can compromise performances. Therefore, the proposed scheme does not employ any model of human faces (or any training method). Although a preliminary detection of human faces in analyzed images might be used (to avoid unnecessary matches between images without any faces) it is not generally required. In other words, the scheme can additionally retrieve similarly looking fragments (even if they do not show human faces) in images of any contents.

Altogether the scheme is based on the following principles:

(a) The input dataset consists of images containing human faces on unpredictable backgrounds (e.g. captured by surveillance systems).

(b) The scheme identifies pairs of images containing fragments which look similar (and approximately localizes such fragments). It is expected that the majority of these fragments are similar faces or their fragments, but similarly looking background fragments can be retrieved as well (e.g. in images are captured on the same or similar backgrounds).

(c) Additionally, the retrieved images can be clustered into groups of images presumably showing the same individuals.

In Section II, we briefly discuss background works on face detection and recognition, and overview tools used in the proposed scheme. Section III (the main part of the paper) describes the scheme and discusses its performances on the exemplary datasets.

Section IV contains the additional results on image clustering, while the concluding remarks and observations are given in Section V.

II. BACKGROUND WORK

A. Automatic face recognition

The main advantages of vision-based face recognition are low costs and convenience. Unfortunately, truly convenient systems (where images are captured by unobtrusively located

cameras) usually have limited performances because of partial occlusions, diversified facial expressions, hairstyles, glasses, etc. Under such conditions, performances of industrial face recognition systems are not very impressive (e.g. 54% hit rate of FaceVASC reported in [5] for images of celebrities collected from internet).

Another factor limiting applicability of fully automatic face identification systems is the complexity of underlying algorithms (see [25]). The most typical mathematical models, like *eigenfaces* (e.g. [14], [21]), EBGM (elastic bunch graph matching, [23]) or LDA (linear discriminant analysis, e.g. [1]) require complex processing operations which might be prohibitively time-consuming for large dataset.

Moreover, these methods assume some knowledge of the human face anatomy (learnt from sample images or provided by human operators, e.g. [5]).

Therefore, even though systems better than humans identifying faces (at least for front-view benchmark images) are known, e.g. [10] and [13], face identification by human observers is still considered more flexible and robust. Nevertheless, preliminary retrieval of faces possibly showing the same individuals is always a welcome tool in face identification over large collections of images.

The proposed scheme assumes no particular knowledge about properties of the face images. We use a general idea of keypoint detection and matching.

Keypoint matching has been used in several works on face identification (e.g. [2]). However, only matches between individual keypoints are generally used, and the proposed algorithms require training (e.g. selection of most representative keypoints in [2]).

B. Keypoint detection, description and matching

Views of human faces are almost always (even in frontal view images) subject to perspective distortions, which are typically locally approximated by affine transformations. Therefore, we use affine-invariant keypoint detectors, namely Harris-Affine and Hessian-Affine, see [7]. Keypoints are represented by SIFT, [6], which is apparently the most popular descriptor.

Keypoint descriptors are usually quantized into finite numbers of visual words. Vocabularies of diversified sizes are used (e.g. [8], [9]) but we propose a vocabulary of 2,000 words only. The discriminative power of such a small vocabulary is very low, but we actually build Cartesian products of small vocabularies (see Section II.D) providing sufficient distinctiveness of the resulting vocabulary.

Typical keypoint matching techniques employ either *mutual-nearest-neighbor* approach (this is a *one-to-one* scheme providing relatively high credibility, but computationally very expensive). Instead, the alternative *the-same-word* approach is universally applied. It is a cheap *many-to-many* scheme, which can flexibly control (depending on the size of vocabulary) the level of *precision* and *recall* (*precision* is proportional to the vocabulary size and *recall* inversely proportional).

C. Augmented keypoint descriptors

Regardless the matching scheme (and regardless the vocabulary size in M2M schemes) credibility of individual keypoint correspondences is very low. Statistics provided in [17], [18] suggest that only 1-2% of keypoint correspondences between unpredictable images indicate actually similar image fragments. Therefore, practically all *state-of-the-art* methods of keypoint-based image retrieval incorporate into the matching process verification of configuration consistency, e.g. [3], [4], [12], [20]. Eventually, only keypoints satisfying the configuration constraints are considered true correspondences. In spite of numerous improvements (especially regarding preliminary retrieval of the most promising keypoints) such verification is a tedious process and the methods employing this process are not fully scalable.

In this paper, we apply an alternative approach based on the principles outlined in [17] and [18]. The basic idea is illustrated in Fig. 1. Given three affine-invariant keypoints K_0, K_1 and K_2 with their corresponding ellipses E_0, E_1 and E_2, three trapezoids Q_0, Q_1 and Q_2 can be formed (their shapes are uniquely defined by the shapes of the ellipses and their relative locations, details are available in [17] and [18]).

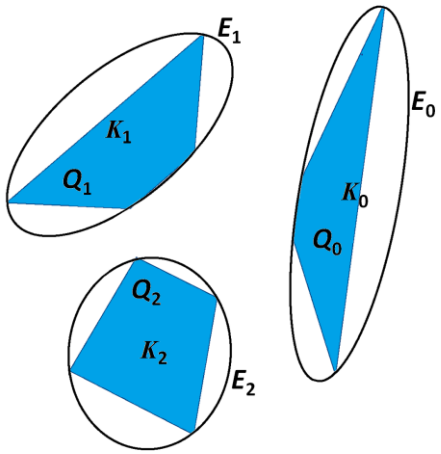

Figure 1. Trapezoids built within a triplet of affine-invariant (elliptic) keypoints.

When a triplet of elliptic keypoints is transformed by an affine mapping, the configuration of Q trapezoids is transformed by the same mapping. Therefore, the affine covariance between such triplets of keypoints can be identified by the same (similar) values of selected affine-invariant shape descriptors calculated over the trapezoids.

Following [17], we apply the simplest affine moment invariant (e.g. [15])

$$Inv = \frac{M_{20}M_{02} - M_{11}^2}{M_{00}^4} \quad (1)$$

where M_{pq} indicates a central moment of $(p + q)^{th}$ order.

Thus, the triplet of K_0, K_1, K_2 keypoints can be affine-invariantly represented by a 3D TERM descriptor (the name proposed in [17]):

$$TERM_{0,1,2} = \left[Inv(Q_0), Inv(Q_1), Inv(Q_2) \right] \quad (2)$$

Actually, in [18] more shapes are considered within the triplets of keypoints so that the descriptor's dimensionality is much higher. We reduce it to only three dimensions in order to tolerate more geometric distortions of the keypoint shapes and locations. For example, the keypoint ellipses can be individually scaled, but the descriptor remains the same if the locations of the keypoints are not changed (while the values of descriptors proposed in [18] would change in such a scenario).

Obviously, the visual content of the triplet can be represented by three SIFT descriptors/words $SIFT(K_0)$, $SIFT(K_1)$ and $SIFT(K_2)$.

Subsequently, description of keypoints in the context of their neighborhoods is created from combinations of *SIFT* and *TERM* descriptors computed over such triplets formed within the keypoint neighborhood.

First, a limited-size neighborhood is built for each extracted keypoint K_0. The neighborhood consists of not more than N (we propose $N = 20$) other keypoints of similar size (for example, between 50% to 150% of the K_0 keypoint area). Moreover, the keypoints selected for the neighborhood must be within a limited distance from K_0 (we propose the range between 70% and 200% of the Mahalanobis distance defined by the size of E_0 ellipse).

Then, using K_0 keypoint and its neighbors $\{K_1, K_2, ..., K_N\}$, we can build a number of $\{K_0, K_i, K_j\}$ triplets. Afterward, K_0 keypoint and its context (neighborhood) are represented by $SIFT(K_0)$ and a $ST(K_0)$ sentence, which is a union of 3-element tuples:

$$ST(K_0) = \bigcup_{i,j} \left\{ SIFT(K_i), SIFT(K_j), TERM_{0,i,j} \right\} \quad (3)$$

The most typical numbers of phrases within such descriptions are between 50 and 80. The numbers are limited (compared to the expected numbers of triplets in a neighborhood of up to 20 keypoints) because triplets forming too narrow triangles are excluded for numerical stability of the calculations. Moreover, there are also keypoints with no such descriptions at all. For example, large keypoints surrounded by only very small keypoints will have no neighboring keypoints suitable for building the triplets.

D. Matching augmented keypoint descriptors

It was mentioned in Section II.B that *SIFT* descriptors are quantized into a 2000-word vocabulary. *TERM* descriptors are also quantized, and the quantization is very coarse. The values of *Inv* invariant (see Eq. 1) are quantized into 12 bins only so that the resulting size of *TERM* vocabulary is $12^3 = 1728$ words.

Altogether, a keypoint with its neighborhood is eventually described by a word from a vocabulary of 2,000 words, i.e. $SIFT(K_0)$, and a number of words (forming $ST(K_0)$ sentence, see Eq. 3) from a very large vocabulary of $2{,}000 \times 2{,}000 \times 1{,}728 = 6{,}912{,}000{,}000$ words.

Therefore, in spite of small sizes of *SIFT* and *TERM* vocabularies, the visual and geometric properties of keypoint neighborhoods are represented by sufficiently large numbers of words which provide distinctive representations of diversified image contents (while tolerating wide margins of photometric and geometric distortions).

Using the above descriptions of keypoints and their contexts (neighborhoods), the keypoint matching process is straightforward. Two keypoints K_0 and L_0 match (i.e. their neighborhoods match as well) if

$$SIFT(K_0) = SIFT(L_0) \quad \text{and} \quad ST(K_0) \cap ST(L_0) \neq \varnothing \quad (4)$$

i.e. the keypoints are visually similar and their neighborhoods are (at least partially) similar both photometrically and geometrically.

Altogether, the keypoint description and matching algorithm follows partially the concept of keypoint bundling (e.g. [24]). However, in other works keypoint bundling is primarily used as a tool for reducing the number of keypoints for the analysis of configuration consistency. We use this concept for a direct keypoint matching.

III. Detection of Similar Faces

The proposed scheme for detection of similar faces in random images is directly based on the abovementioned method is keypoint matching. Two images are matched (i.e. they presumably contain similarly looking faces) if there is at least one keypoint correspondence (defined by Eq. 4) found between the images.

The scheme obviously does not take into account any particular characteristics of human faces. Therefore, a pair of images containing any similarly looking components would be matched as well. However, we assume that datasets of processed images contain primarily human faces captured on diversified backgrounds so that detection of accidental similarities between the backgrounds is accepted as unavoidable "collateral damage".

Because the proposed keypoint descriptions tolerate relatively wide ranges of photometric and geometric distortions, there is always a danger that the matched image fragments are insufficiently similar (but the corresponding pairs of images are, nevertheless, retrieved). To minimize such effects, two keypoint detectors are independently used. Harris-Affine (HaA) extracts primarily corner-related keypoints, while Hessian-Affine (HeA) returns keypoint corresponding to blobs. In order to accept a semi-local similarity between images, both types of matched keypoint should exist in conjunction.

Therefore, a pair of matched HaA keypoints (K_{HA}, L_{HA}) is eventually accepted only if there is another pair of HeA keypoints (K_{HE}, L_{HE}) overlapping it (and another way around). "Overlapping" means that the Mahalanobis distances (defined by the sizes of relevant ellipses) between the corresponding keypoints are below the threshold. Fig. 2 provides an example.

The above scheme has been preliminarily tested on two popular datasets:

(a) Caltech faces[1] (note that we use the difficult part of the dataset, i.e. faces on wider backgrounds).

(b) Georgia Tech faces[2] (in contrast to the previous dataset we use cropped images containing only the faces).

For the sake of uniformity, each face is represented by 15 images in both datasets (even though some faces in Caltech dataset are shown on more images).

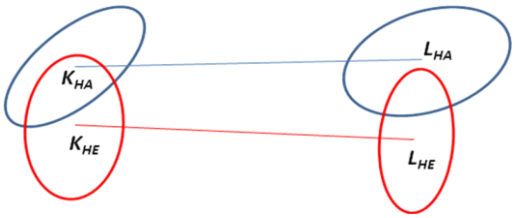

Figure 2. A pair of matched HaA keypoints (blue) overlapping a pair of matched HeA keypoints (red). The threshold value is $\sqrt{2}$.

A. *Caltech faces experiment*

The processed part of this dataset contains 19 faces represented by 15 images each. All pairs of images are compared so that the total number of matched image pairs is 40,470. The number of *ground truth* matches (i.e. pairs of images showing the same face) is 1,995.

By using the proposed scheme with Harris-Affine keypoint detector, the algorithm retrieves 6,176 image pairs. When Hessian-Affine is applied, the number of retrieved pairs is very similar, i.e. 6,035.

In both variants, a significant number of ground truth pairs has been retrieved (1,636 and 1,582, correspondingly) so that *recall* is acceptably high at 82% and 79.3% levels. Similar results have been obtained in [16] by using much more advanced (and time consuming) alternative techniques based on keypoint matching. Nevertheless, *precision* of the results is unacceptably low (26.5% and 26.25%, correspondingly) even if some of false positives are actually correct because they indicate not the same faces but identical background fragments. (see examples in Fig. 4).

Much better performances (in terms of *precision*) are obtained by "overlapping" HaA and HeA results. The total number of retrieved image pairs is then reduced to 1,876, of which 1,472 are *ground truth* pairs. Examples are provided in Fig. 3. Therefore, a satisfactory 73.8% *recall* is combined with equally satisfactory 78.5% *precision*. Nevertheless, the actual *precision* is even higher (at approx. 92% level) because about 250 retrieved image pairs contain identical background fragments, which are detected instead of the same faces. Examples of such cases are provided in Fig. 4.

Figure 3. Examples of correct face retrieval in Caltech dataset. Note that each keypoint match actuallt indicates an unspecified number of matches within the corresponding neighborhoods.

Figure 4. Examples of correctly retrieved similarly looking background fragments (even though the images contain different faces).

Only approx. 100 retrieved pairs are "true *false positives*". They actually represent similarities between fragments of otherwise different faces (e.g. similar hair, eyes, chins, etc.). Selected examples are given in Fig. 5, but more detailed analysis of this issue is available in [16] and [19].

[1] http://www.vision.caltech.edu/Image_Datasets/faces/faces.tar
[2] http://www.anefian.com/research/GTdb_crop.zip

B. Georgia Tech faces experiment

In this experiment, we used 22 faces, each represented by 15 images. Unlike in the previous test (where only near-frontal views are available) these images show highly diversified views, including different expressions and significant viewpoint changes. However, the image backgrounds are almost entirely cropped so that very few (if any) matches between background areas can be expected.

Figure 5. True *false positives* (in Caltech dataset) which, nevertheless, represent similarities between fragments of different faces.

There are 54,285 image pairs altogether and 2,310 of them are considered *ground truth* (pairs showing the same face). In contrast to the previous test, the numbers of image pairs retrieved by using only HaA or HeA detector are not very high, i.e. 1,411 and 1,592 correspondingly.

The numbers of *true positives* are, correspondingly, 1,070 and 1,076. The *recall* is, therefore, slightly below 50% which is understandable for faces looking sometimes very differently (examples in Fig. 6). However, *precision* is quite satisfactory (75.8% and 67.6%).

Figure 6. Examples of image diversity within *the same face* category.

Thus, we can preliminarily claim that a large number of *false positives* retrieved in Caltech dataset by using individual keypoint detectors is caused primarily by accidental similarities between image backgrounds. When backgrounds are practically non-existing, the need for "overlapping" results is reduced. Nevertheless, the combination of HaA and HeA results provides even better performances. The number of retrieved *true positives* is 995 (i.e. there is only a marginal drop of *recall*) while *precision* improves to 94.4% because only 1,054 images pair are retrieved altogether by the combination of HaA and HeA. Exemplary correct retrievals are given in Fig. 7, while some (from very few ones) *false positives* are shown in Fig. 8.

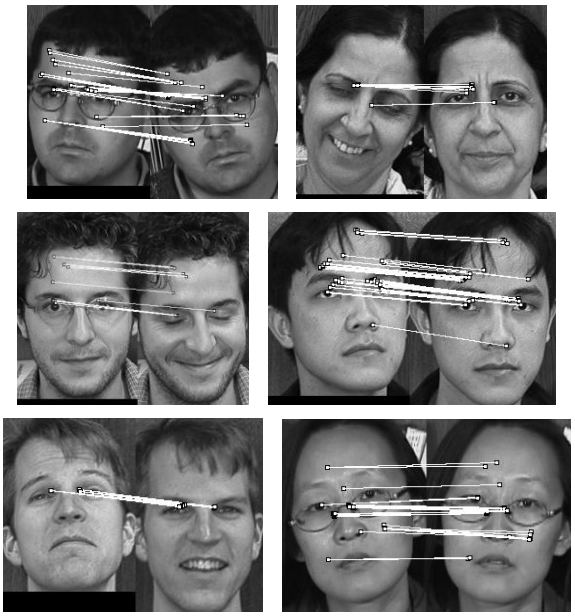

Figure 7. Exemplary correct rertievals from Georgia Tech dataset.

Thus, we can preliminarily claim that a large number of *false positives* retrieved in Caltech dataset by using individual keypoint detectors is caused primarily by accidental similarities between image backgrounds. When backgrounds are practically non-existing, the need for "overlapping" results is reduced. Nevertheless, the combination of HaA and HeA results provides even better performances. The number of retrieved *true positives* is 995 (i.e. there is only a marginal drop of *recall*) while *precision* improves to 94.4% because only 1,054 images pair are retrieved altogether by the combination of HaA and HeA. Exemplary correct retrievals are given in Fig. 7, while some (from very few ones) *false positives* are shown in Fig. 8.

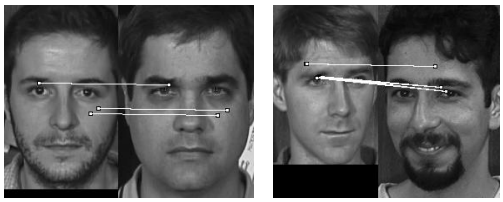

Figure 8. Exemplary incorrect rertievals from Georgia Tech dataset.

Of course, none of the results presented in Sections III.A and III.B can be compared to performances reported for specialized *state-of-the-art* face recognition systems. The best algorithms overviewed in [13] provide.90% *recall* with 99% *precision* on high-resolution frontal view images under uncontrolled illuminations. It should be noted, nevertheless, that the proposed scheme is straightforward and extremely efficient. Only individual keypoint matches are used (and a simple "overlap" of HaA and HeA results) so that overall complexity is $O(N \cdot M)$, where N and M are the corresponding numbers of keypoints in a pair of matched images.

A scheme of such a complexity is scalable to databases of very large sizes. Moreover, the scheme can be adopted without any changes to images from other domains. When matching is applied to sequences of images (e.g. video-frames from surveillance cameras) rather than to individual images, the expected performances can be statistically almost as good as for those *state-of-the-art* systems. This is because the statistically expected *recall* for a sequence of 3-4 frames showing the same face and matched against a single image is at 90-95% level (with the correspondingly high *precision*, as discussed above).

IV. Clusters of Face Images

In practice, a dataset of matched image may consist of images with totally unpredictable contents, and one of the additional objectives can be to identify groups of images showing the same faces (or showing other the same objects, which is an unwelcome, but a possible outcome).

For that purpose, a *similarity graph* can be straightforwardly built for the dataset images. Nodes of the graph represent images, and two nodes are linked if that pair of images is retrieved by the matching scheme. Theoretically, nodes of fully connected sub-graphs within such a similarity graph would represent groups of images sharing the same face (or another identical object).

However, as shown in the conducted tests, not all pairs of images with the same face(s) are retrieved, and some retrieved image pairs show different faces. Therefore, we propose to cluster nodes of the similarity graph using *k*-connected sub-graphs (a similar idea has been used in [11]).

The above concept was tested on the Caltech and Georgia Tech results discussed in Section III. The results are sound, though not fully satisfactory. First, we have identified that 3-*connectivity* is the variant of *k*-connectivity recommended for this aplication. In Georgia Tech dataset, 20 sub-graphs (representing 20 faces out of 22 faces present in the dataset) have been formed. Two faces in this dataset are so diversified that their images do not form any 3-connected graph.

However, only two of the sub-graphs (image clusters) include all fifteen images showing the same face. Other clusters are smaller, and Fig. 9 shows an exemplary cluster of images identified as *the-same-face* images (note that it contains only 8 out of 15 database images actually showing this face).

In the Caltech dataset, however, there are many pairs of images sharing identical background fragments. Therefore, we have identified cases where two (or more) different faces are included into the same 3-connected sub-graph because the images are linked by the same background fragments (rather than by the same faces). Fig.10 shows an illustrative example of such a cluster.

Figure 9. A cluster of *the-same-face* images automatcially found in Georgia Tech dataset.

Figure 10. Fragments of an incorrect cluster of *the-same-face* images automatically found in Caltech dataset.

V. SUMMARY

The paper investigates performances of a simple keypoint-based scheme for retrieval pairs of images containing the same face. The scheme is computationally very efficient and, therefore, can be applied to large visual datasets.

The main components of the scheme are:

(a) Two independent affine-invariant keypoint detectors, i.e. Harris-Affine and Hessian-Affine. The results obtained by using both detectors are "overlapped" in the final phase of the scheme, and only the results with approximately the same locations of keypoints matched by both variants are retained.

(b) A novel keypoint detector is used. The detector represents visual properties of the keypoint itself and, additionally, visual and geometric properties of the limited-size keypoint neighborhood (context). Having such a descriptors, similar image fragments (in a wider context than similarity between individual keypoints) can be identified by straightforward correspondences between keypoints.

(c) Visual properties of keypoints are represented by *SIFT* descriptor (quantized into a 2,000-word vocabulary) while the neighborhood geometry is affine-invariantly represented by a recently proposed *TERM* descriptor (which is also quantized into a 1,728-word vocabulary). Altogether, keypoints are described by a set of phrases obtained as the Cartesian product of these vocabularies. In this way, the matching results are insensitive to a wide range of photometric and geometric distortions (which are typically present in face images due to hairstyles, glasses, expressions, etc.) while retaining the most significant structures in the images.

The scheme does not require any training or knowledge about anatomy of the human faces. Therefore, it can be directly applied to other problems with similar requirements (e.g. detecting dogs of the same breed, similar flowers, etc.). Nevertheless, this is also a disadvantage because the scheme does not distinguish between images containing the same faces and images containing other similarly looking components. Therefore, a preliminary detection of images containing human faces (using one of available algorithms, e.g. [22]) would be a recommended pre-retrieval module for the scheme.

Because of the above disadvantage, the scheme cannot fully identify clusters of images containing the same faces on unpredictable background (which would be a step towards automatic annotation) since accidental similarities between background components can link images containing different faces.

The paper does not discuss the implementation issues of the proposed scheme. It is obvious, nevertheless, that an algorithm of such a low complexity can be prospectively used with databases of very large sizes and/or in real-time applications.

REFERENCES

[1] Belhumeur, P.N., Hespanha, J.P. and Kriegman, D.J., 1997, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. PAMI 19(7), 711-720.

[2] Bicego,M., Grosso, E., Lagorio, A., Brelstaff, G., Brodo, L. and Tistarelli, M., 2008, Distinctiveness of Faces: A Computational Approach, ACM Trans. Applied Perception 5(2), 11.1-11.17.

[3] Chum, O., Perdoch, M. and Matas, J., 2009, Geometric min-hashing: Finding a (thick) needle in a haystack, Proc. IEEE Conf. CVPR'09, 17–24.

[4] Jegou, H., Douze, M. and Schmid, C., 2010, Improving bag-of-features for large scale image search Int. Journal of Computer Vision 87(3), 316–336.

[5] Jenkins, R. and Burton, A.M., 2008, 100% Accuracy in Automatic Face Recognition, Science 319, 435.

[6] Lowe, D., 2004, Distinctive Image Features from Scale-invariant Keypoints, Int. Journal of Computer Vision 60(1), 91–110.

[7] Mikolajczyk, K. and Schmid, C., 2004, Scale and Affine Invariant Interest Point Detectors, Int. Journal of Computer Vision 60(1), 63–86.

[8] Mikulik, A., Perdoch, M., Chum, O. and Matas, J., 2012, Learning vocabularies over a fine quantization. Int. Journal of Computer Vision, doi: 10.1007/s11263-012-0600-1.

[9] Nister, D. and Stewenius, H., 2006, Scalable recognition with a vocabulary tree, Proc. IEEE Conf. CVPR'06, 2161-2168.

[10] O'Toole, A.J., Phillips, P.J., Jiang, F., Ayyad, J., Penard, N. and Abdi, H., 2007, Face recognition Algorithms Surpass Humans Matching Faces across Changes in Illumination, IEEE Trans. PAMI 29(9), 1642-1646.

[11] Paradowski, M. and Sluzek, A., 2010, Automatic Visual Object Formation using Image Fragment Matching**,** 5 Int. Symp. AAIA 2010, 97-104.

[12] Paradowski, M. and Sluzek, A., 2011, Local keypoints and global affine geometry: triangles and ellipses for image fragment matching, in: Innovations in Intelligent Image Analysis (eds. H.Kwasnicka, L.Jain), Springer Verlag, Vol. SCI339, 195-224.

[13] Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L. and Sharpe, M., 2007, FRVT 2006 and ICE 2006 Large-Scale Results, National Institute of Standards and Technology Techn. Report NISTIR 7408.

[14] Sirovich, L. and Kirby, M., 1987, Low-dimensional Procedure for the Characterization of Human Faces, Journal of the Optical Society of America A 4(3), 519–524.

[15] Sluzek, A., 1990, Zastosowanie metod momentowych do identyfikacji obiektow w cyfrowych systemach wizyjnych. WPW, Warszawa.

[16] Sluzek, A. and Paradowski, M., 2012, Visual Similarity Issues in Face Recognition, Int. Journal of Biometrics (1), 22-37.

[17] Sluzek, A. and Paradowski, M., 2012, Detection of Near-duplicate Patches in Random Images using Keypoint-based Features, Proc. ACIVS 2012, LNCS 7517, 301-312.

[18] Sluzek, A., 2012, Large Vocabularies for Keypoint-Based Representation and Matching of Image Patches, Proc. ECCV 2012 W&T, LNCS 7583, 229-238.

[19] Sluzek, A., Paradowski, M. and Yang, D., 2012, Reinforcement of Keypoint Matching by Co-segmentation in Object Retrieval: Face Recognition Case Study, Proc. ICONIP 2012, LNCS 7667, 34-41.

[20] Stewenius, H., Gunderson, S.H. and Pilet, J., 2012, Size matters: Exhaustive geometric verification for image retrieval, Proc. ECCV 2012, LNCS 7573, 674-687.

[21] Turk, M. and Pentland, A., 1991, Eigenfaces for Recognition, Journal of Cognitive Neuroscience 3(1), 71–86.

[22] Viola, P. and Jones, M., 2004, Robust Real-time Face Detection, Int. Journal of Computer Vision 57(2), 137-154.

[23] Wiskott, L., Fellous, J.-M., Kruger, N. and von der Malsburg, Ch., 1999, Face Recognition by Elastic Bunch Graph Matching, in: Intelligent Biometric Techniques in Fingerprint and Face Recognition (eds. L.C. Jain *et al.*), CRC Press, 355-396.

[24] Wu, Zh., Ke, Q., Isard, M. and Sun, J., 2009, Bundling features for large scale partial-duplicate web image search. Proc. IEEE Conf. CVPR'09, 25–32, 2009.

[25] Zhao, W., Chellappa, R., Phillips, P.J. and Rosenfeld, A., 2003, Face Recognition: A Literature Survey, ACM Computer Surveys 35, 399–458.