# Scalable Face Retrieval by Simple Classifiers and Voting Scheme

Yuzuko Utsumi(✉), Yuji Sakano, Keisuke Maekawa, Masakazu Iwamura, and Koichi Kise

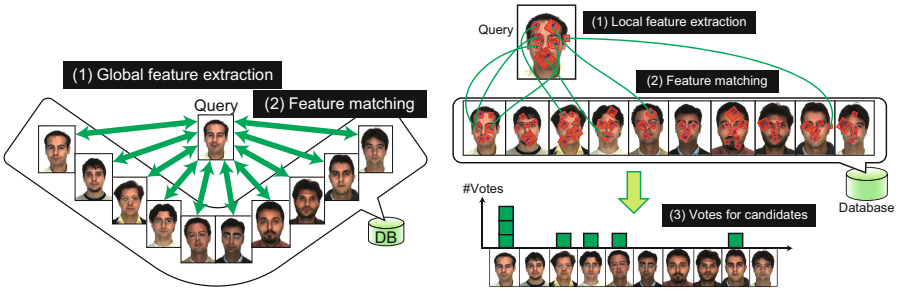
Graduate School of Engineering, Osaka Prefecture University,
1-1 Gakuencho, Naka, Sakai, Osaka 599-8531, Japan
{yuzuko,masa,kise}@cs.osakafu-u.ac.jp

**Abstract.** In this paper, we propose a scalable face retrieval method on large data. In order to search a particular person from videos on the Web, face recognition is one of the most effective methods. Needless to say that retrieving faces from videos are more challenging than that from a still image due to inconsistency in imaging conditions such as change of view point, lighting condition and resolution. However, dealing with them is not enough to realize the retrieval on large data. In addition, a face recognition method on the videos should be highly scalable as the number of the videos on the Web is enormous. Existing face recognition methods do not scale. In order to realize scalable face recognition, we propose a novel face recognition method. The proposed method is scalable even if the data is million-scale with high accuracy. The proposed method uses local features for face representation, and an approximate nearest neighbor (ANN) search for feature matching to reduce computational time. A voting scheme is used for recognition to compensate for low accuracy of the ANN search. We created a 5 million database and evaluated the proposed method. As results, the proposed method showed more than thousand times faster than a conventional sublinear method. Moreover, the proposed method recognized face images with a top 1000 cumulative accuracy of 100% in 139 ms recognition time (excluding preprocessing and feature extraction for the query image) per query image on the 5 million face database.

## 1 Introduction

Imagine that you want to find an actress from videos on the Web. It is hard to find her manually because the number of videos on the Web is enormous. Therefore, retrieving particular persons by a machine from enormous video data is demanded. Face recognition is useful for finding a particular person. In order to realize the retrieval, we need a face recognition method which is robust against inconsistency in imaging conditions such as change of view point, lighting condition and resolution. However, dealing with them is not enough to realize the retrieval on large data. In addition, scalability is required because face recognition on the video, meaning face recognition on large data, requires a lot of computational time. In this paper, we focus on the scalability.


Q. Ji et al. (Eds.): FFER 2014, LNCS 8912, pp. 99–108, 2015.
DOI: 10.1007/978-3-319-13737-7_9




(a) A typical conventional face recognition method

(b) The proposed method

**Fig. 1.** Overview of face recognition processes

Scalable face recognition on large data is not easy. Most conventional methods are based on 2-class classifiers, e.g., [1,7,10,22,25,27,29,30,32]. Figure 1a shows an example of the conventional face recognition methods. When the methods recognize a face, (1) a feature is extracted from a query image and (2) the method must evaluate the query as many times as the number of persons (classes) the method can recognize. As a result, those methods take computational time in proportion to the size of data. Exceptionally, some methods take into account scalability [24,26]. Those methods can recognize a face with sublinear time to the size of data. However, the sublinear time just means less than linear time, and does not always mean scalable enough; computational times of methods in sublinear times can vary from thousands to tens of thousands.

Therefore, we propose a scalable sublinear face recognition method to retrieve face images from a large-scale data. As shown in Fig. 1b, the proposed method represents a query image as multiple local features and recognizes face images by a voting scheme with a classifier based on an approximate nearest neighbor (ANN; approximate NN) search method. Thanks to the ANN, the proposed method can reduce computational time substantially. Meanwhile, the ANN search makes a sacrifice of matching accuracy. The voting scheme compensates for low matching accuracy. We created a 5 million database for evaluating scalability of the proposed method. As results, the proposed method was thousand times faster than the conventional sublinear method [26].

The proposed method can be applied to a face sequence extracted from a video. This means that we can use multiple images for recognition. As shown in [8], face recognition using image sequence shows better recognition rate than that of using a single image. Therefore, we can execute more accurate face retrieval when the proposed method is used for video data.

## 2 Related Work

Face recognition is an active research area and various methods have been proposed. Conventional methods like Eigenface [25,27,29], Fisherface [1,30], Linear

Discriminant Analysis (LDA) [10,32] and Support Vector Machine (SVM) [7,22] are not useful on large data because their computational times for recognition increase linearly. Their weak point is that the classifiers are designed to solve 2-class distinction problems. In order to recognize a face from multiple persons, the 2-class classifiers have to be structured using one-vs-one or one-vs-all approach. The computational time of the one-vs-all approach, which is lighter than the one-vs-one approach in computational cost, increases in proportion to the number of categories in the database. Thus it is very time consuming and they are impractical on a large-scale database.

Some face recognition methods realized sublinear computational time through decreasing computational time of classifiers. Schwartz et al. [24] proposed a scalable face recognition method based on a decision tree. Shi et al. [26] proposed a rapid face recognition method, which speeds up a sparse representation method [28] by up to 150 times by using a hashing method. They succeed in decreasing computational costs, but they still have problems with scalability because they are evaluated with small size data. There is no confidence that they are scalable on large data.

In object retrieval, some scalable methods based on the Bag-of-Features (BoF) model have already proposed [3,14,20]. However, they are not designed for specific object recognition like face recognition but for generic object recognition. In the BoF model, local features extracted from an image are quantized by visual words, and the image is represented by a histogram of the visual words. They can realize scalable object retrieval because the BoF model can express images more compact than local features. However, quantization is known to decrease discriminative power; it is shown by Zhu et al. in the context of visual object retrieval that local feature matching without quantization outperformed the BoF model (local feature matching with quantization) [33]. Therefore, for retrieving face images we use the strategy of local feature matching without quantization, following the success of the strategy in the specific object recognition [11,12,17].

## 3 Proposed Method

The proposed method consists of face representation by multiple feature vectors per image and a voting scheme with NN classifiers using ANN search instead of NN search. The classifiers are quite fast while recognition rate is not high due to the trade-off relationship between accuracy and computational time. Thus it is hard to avoid reduction of recognition accuracy for fast classifiers. However, the voting scheme compensates for the reduction. With the voting scheme, even if the classifier performs moderate recognition accuracy, a voting result can achieve high recognition accuracy. This is because the true class tends to have more votes than others even if casting each vote is less accurate. In order to utilize the voting scheme, multiple feature vectors are required to be extracted from a query image. This is a reason to use a local feature.

Detailed procedure of the proposed method is presented. As preparation, local features extracted from reference images are stored in the database with

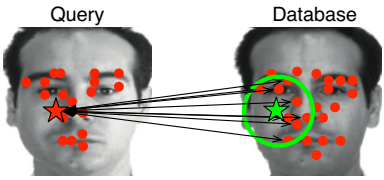



**Fig. 2.** Example of position constraint for feature matching. The green circle in the face image in the database (on the right side) represents the position constraint for the feature vector represented by a red star in the query face image (on the left side).

class labels (i.e., IDs of reference images). For a given query face image, the following process is carried out (the numbers correspond to those of Fig. 1b). (1) features are also extracted from the query image. (2) in each feature of a query image, K nearest features in the database are calculated by the ANN search method in the Euclidean distance. (3) votes are cast for the class labels of the retrieved K nearest features. Consequently, the class label which gets the maximum votes is chosen as the recognition result. We use a weighted voting scheme. The weight is determined based on a distance between the nearest feature to the query feature in the feature space. By letting $d$ be the distance, the weight of the voting of the matched feature is represented as $1/d^2$.

We use geometric limitation to feature matching as shown in Fig. 2. That is, we limit the area of features in the database for matching a query feature. The feature matching method can avoid mismatching features extracted from different part of faces, though there is a chance to fail in recognizing misaligned face images as recent methods such as [2] do.

## 4 Experiments

### 4.1 Experimental Environment

For evaluation, we created a 5 million face database, which consisted of face images downloaded from the Web, Flickr[1], and public face databases: AR Face Database [16], the Extended Yale B Face Database [5], CAS-PEAL [4], FERET [21], The AT&T The Database of Faces[2], Georgia Tech Face Database [19], Surveillance Cameras Face Database [6] and Indian Face Database[3]. The samples of the data from the Web are shown in Fig. 3. We used some keywords and dates to search, and downloaded the images. After downloading the images, we got rid of duplicate image files. Therefore, the images from the Web may contain multiple images of one person. The face images were cropped by using a face detector [18] and face direction of the cropped images was normalized by using the method proposed in [13,31], which extracted 14 facial feature points on eyes, nose and so on.

[1] Flickr, https://www.flickr.com
[2] AT&T The Database of Faces, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
[3] The indian face database, http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/

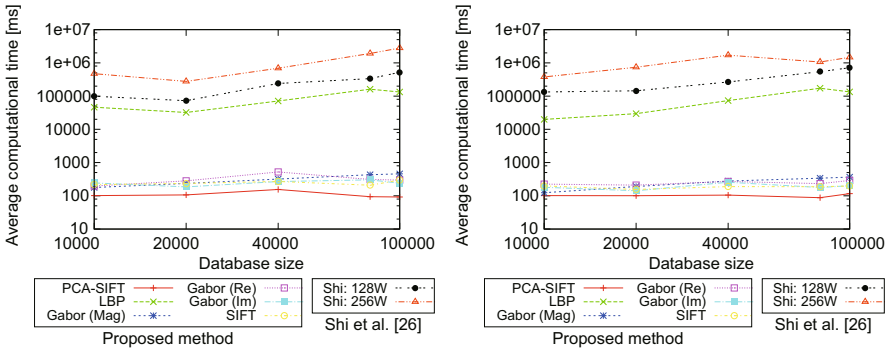
**Fig. 3.** Samples of face images downloaded from the Web

Based on these facial feature points, we extracted a face region from a face image and normalized face direction. Normalized images were 8 bit gray-scale and the dimensions were $512 \times 512$ pixels.

We used the AR Face Database [16] and the Extended Yale Face Database B [5] as test data. The AR Face Database consists of over 4,000 images from 136 individuals. We put the set 1 (Neutral) to the database. We excluded two images of the set 1 from the database because they failed to be normalized. We put the sets 2-7 (Smile, Anger, Scream, Left Side Light on, Right Side Light on, and All Side Light on) to test sets (query images). The number of queries was 792. The Extended Yale Face Database B consisted of 16,128 images from 28 individuals. We put 27 images from 27 individuals whose facial pose was frontal and light source direction with respect to the camera axis was at 0 degrees azimuth and 0 degrees elevation to the database. We also excluded one image from the database because the detection or normalization failed. We used the remaining images, where the detection and normalization succeeded, as queries. Images of the test set were also cropped and normalized by the same method as the database. The number of queries was 847. In the following results, computational time excludes the time for preprocessing and feature extraction for the query image. We employed a computer with AMD Opteron 2.2GHz CPU and 256GB RAM. The average number of feature vectors extracted from an image was about 200. We used the state-of-the-art ANN search method that is empirically shown to have the computational time less than proportional (sublinear) to the data size [23].

### 4.2 Experimental Results

In order to compare the recognition accuracy and recognition speed, we evaluated the proposed method and a sublinear method proposed in [26]. In the proposed method, we used PCA-SIFT [9], SIFT [15], the Gabor wavelet feature (real part, imaginary part and magnitude), and the Local Binary Patterns as local features. The dimensionality of PCA-SIFT, SIFT, the Gabor wavelet feature, and the LBP were 36, 128, 40, and 59, respectively. We changed parameters of the K-NN method and the ANN method for evaluation. In the conventional method, we used $256 \times 256$ and $128 \times 128$ images as feature. We also changed the parameters of the conventional method for evaluation. We changed the number of images in a database from 10,000 up to 100,000 and evaluated the scalability of the proposed method and the conventional method. The database we used for




(a) AR Face Database (b) Extended Yale Face Database B

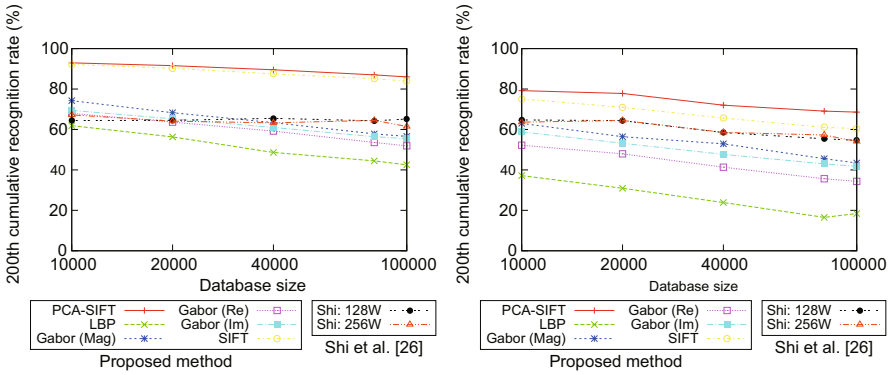
**Fig. 4.** Relationship between the size of the database and computational time

the evaluation is a part of the 5 million database we made and contained AR face database and Extended Yale Face Database B.

The graphs in Figs. 4 and 5 plotted the relationship between the size of the database and computational time, and the relationship between the size of the database and top 200 cumulative recognition rate, respectively. In the graphs, PCA-SIFT, SIFT, Gabor (Mag0, Gabor (Re), Gabor (Im) and LBP show the results of the proposed method with PCA-SIFT, SIFT, magnitude, real part and imaginary part of Gabor wavelet features, and the Local Binary Patterns, and Shi:128 and Shi:256 show the results of the conventional method with $128 \times 128$ pixels and $256 \times 256$ pixels images. From Fig. 4, the proposed method with the PCA-SIFT features was faster than the one with other features. both on AR face database and Extended Yale B database. The proposed method with the PCA-SIFT features was more than thousand times faster than the sublinear method [26]. Moreover, the proposed method did not change computational time when the database size became large. This means that the proposed method realized fast and scalable face recognition. From Fig. 5, the proposed method showed the best recognition rate. This means that the voting scheme compensated for the low matching accuracy, and the proposed method achieved fast and accurate face recognition.

We used the 5 million face database to evaluate the recognition rates and computational time. We used queries from AR Face Database and top 1000 cumulative recognition for an indicator of recognition accuracy. Table 1 shows the recognition rates and computational time in each image set. We achieved 100% recognition rate with 139 ms processing time when the set 5 was used as a test set. The processing time of exhaustive search[4] was 7,685 seconds with 100 % recognition rate and thus the processing time of the proposed method was

[4] The exhaustive k-NN search method we employed had a mechanism to quit distance calculation when it comes out that a feature vector does not have smaller distance than the tentative k-th nearest neighbor.




(a) AR Face Database (b) Extended Yale Face Database B

**Fig. 5.** Relationship between the size of the database and recognition rates. Recognition rate is a top 200 cumulative recognition rate.

**Table 1.** Recognition rates and computational times on the 5 million face database

| Test Sets | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| Recognition Rate [%] | 98.5 | 99.2 | 42.4 | 100 | 97.7 | 39.4 |
| Time [ms] | 156 | 85.9 | 176 | 139 | 128 | 257 |

55,286 times faster than that of exhaustive search. When the Set 4 (Scream) and 7 (All Side Light on) were used for test sets, the recognition rate was lower than other test sets. However, we think that the proposed method is used for video face recognition, the weak point of the proposed method is not serious because we can choose the best facial images from a video sequence or simply accumulate features or result in each image.

In order to evaluate the scalability of the proposed method, we calculated the average computational time at different numbers of images in the database: 200 thousand, 500 thousand, 1 million, 3 million and 5 million. We used the Set 5 for the test set. We adopted the shortest computational time in the condition that the recognition rate was higher than 98%. Fig. 6 shows the experimental results. From this results, the processing time increased with increasing the number of images in the database monotonically. However, the gradient from 3 million images to 5 million images was gentler than that from 500 thousand images to 3 million images. Increase of processing time of the proposed method was obviously less than proportional to the number of images in the database. Thus, it is indicated that the proposed method has the scalability for a large database.

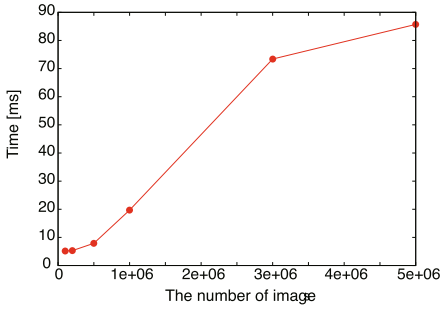



**Fig. 6.** The number of images vs. computational time.

## 5 Conclusion

In this paper, we proposed a scalable face recognition method by using an ANN search method and a voting scheme to achieve fast retrieval on large data in order to find a particular person from videos on the Web. Because of the ANN search method, the proposed method is more scalable than the conventional sublinear method. The ANN search method is fast, while its matching accuracy is low. However, thanks to the voting scheme, the proposed method showed the better recognition accuracy than the conventional sublinear methods. From the experimental results, the proposed method recognized face images with an accuracy of 100% in 139 ms recognition time (excluding preprocessing and feature extraction for the query image) per query image on the 5 million face database when images with an illumination change from left side were used for the query set. The results showed that the proposed method can be applied to video because the proposed method is enough scalable and accurate.

In future work, we evaluate the proposed method with face images cropped from videos on the Web. We can get a face image sequence from a video and use for recognition. If we can use a face image sequence efficiently, we can get better recognition results than the current experimental results.

## Acknowledge

This work was supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society," Funds for integrated promotion of social system reform and research and development of the MEXT, the Japanese Government.

## References

1. Belhumber, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transaction on Pattern Analysis and Machine Intelligence 19(7), 711–720 (1997)

2. Chen, D., Cao, X., Wen, F., Sun, J.: Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In: Proc. 2013 IEEE Conference on Computer Vision and Pattern Recognition (2013)
3. Csurka, G., Dance, C.R., Fan, L., Willamowski, J., Bray, C.: Visual categorization with bags of keypoints. In: Workshop on Statistical Learning in Computer Vision, ECCV, pp. 1–22 (2004)
4. Gao, W., Cao, B., Shan, S., Zhou, D., Zhang, X., Zhao, D., Al, S.: The CAS-PEAL large-scale Chinese face database and evaluation protocols. Joint Research & Development Laboratory, Technical report (2004)
5. Georghiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: Illumination cone models for face recognition under variable lightning and pose. IEEE Transaction on Pattern Analysis and Machine Intelligence **23**(6), 643–660 (2001)
6. Grgic, M., Delac, K., Grgic, S.: Scface - surveillance cameras face database. In: Multiedia Tools and Applications, pp. 1–17 (2009)
7. Heisele, B., Ho, P., Poggio, T.: Face recognition with support vector machines: Global versus component-based approach. In: Proceedings on Eighth IEEE International Conference on Computer Vision, Vancouver, Canada, vol. 2, pp. 688–694 (2001)
8. Iwamura, M., Kunze, K., Kato, Y., Utsumi, Y., Kise, K.: Haven't we met before? – a realistic memory assistance system to remind you of the person in front of you. In: Proc. 5th Augmented Human (AH 2014), March 2014
9. Ke, Y., Sukthankar, R.: PCA-SIFT: A more distinctive representation for local image descriptors. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 506–513 (2004)
10. Kim, T., Stenger, B., Kittler, J., Cipolla, R.: Incremental linear discriminant analysis using sufficient spanning sets and its applications. International Journal of Computer Vision **91**, 216–232 (2010)
11. Kise, K., Noguchi, K., Iwamura, M.: Memory efficient recognition of specific objects with local features. In: Proceedings of the 19th International Conference of Pattern Recognition (2005)
12. Kise, K., Noguchi, K., Iwamura, M.: Robust and efficient recognition of low-quality images by cascaded recognizers with massive local features. In: Proceedings of the 1st International Workshop on Emergent Issues in Large Amount of Visual Data (WS-LAVD 2009), pp. 2125–2132 (2009)
13. Kozakaya, T., Yamaguchi, O.: Face recognition by projection-based 3D normalization and shading subspace orthogonalization. In: FGR 2006, pp. 163–168 (2006)
14. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2169–2178 (2006)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision. **60**, 91–110 (2004)
16. Martínez, A., Benavente, R.: The AR face database. Technical Report 24, Computer Vision Center, Barcelona (1998). http://www2.ece.ohio-state.edu/aleix/ARdatabase.html
17. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transaction on Pattern Analysis and Machine Intelligence **27**(10), 1615–1630 (2005)
18. Mita, T., kaneko, T., Stenger, B., Hori, O.: Discriminative feature co-occurrence selection for object detection. IEEE Transaction on Pattern Analysis and Machine Intelligence 30(7), 1257–1269 (2008)

19. Nefian, A.V., Khosravi, M., III, M.H.H.: Real-time human face detection from uncontrolled environments. In: SPIE Visual Communications on Image Processing (1997)
20. Nowak, E., Jurie, F., Triggs, B.: Sampling Strategies for Bag-of-Features Image Classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 490–503. Springer, Heidelberg (2006)
21. Phillips, J., Moon, H., Rizvi, S.A., Rauss, P.J.: The feret evaluation methodology for face-recognition algorithms. IEEE Trasaction on Pattern Analysis and Machine Intelligence **22**(10), 1090–1104 (2000)
22. Phillips, P.J.: Support vector machines applied to face recognition. Advances in Neural Information Processing Systems **11**, 803–809 (1998)
23. Sato, T., Iwamura, M., Kise, K.: Fast and memory efficient approximate nearest neighbor search with distance estimation based on space indexing. Prmu 2012–142, IEICE Technical Report (February 2013)
24. Schwartz, W.R., Guo, H., Davis, L.S.: A Robust and Scalable Approach to Face Identification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 476–489. Springer, Heidelberg (2010)
25. Sharma, A., Dubey, A., Tripathi, P., Kumar, V.: Pose invariant virtual classifiers from single training image using novel hybrid-eigenfaces. Neurocomputing **73**, 1868–1880 (2010)
26. Shi, Q., Li, H., Shen, C.: Rapid face recognition using hashing. In: Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2753–2760 (2010)
27. Turk, M., Pentland, A.: Eigenface for recognition. Journal of Cognitive Neuroscience **3**(1), 71–86 (1991)
28. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transaction on Pattern Analysis and Machine Intelligence **31**(2), 210–227 (2009)
29. Yang, J., Ahang, D., Frangi, A.F., yu Yang, J.: Two-dimensional pca: a new approach to appearance-based face representation and recognition. IEEE Transaction on Pattern Analysis and Machine Intelligence 26(1), 131–137 (2004)
30. Yang, M.H.: Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In: Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 215–220. Washington, DC, May 2002
31. Yuasa, M., Kozakaya, T., Yamaguchi, O.: An efficient 3D geometrical consistency criterion for detection of a set of facial feature points. IEICE - Trans. Inf. Syst. E91-D, 1871–1877 (2008)
32. Zhao, W., Chellappa, R., Krishnaswamy, A.: Discriminant analysis of principal components for face recognition. In: Proceedings on IEEE International Conference on Face and Gesture Recognition (FG1998), pp. 14–16. Nara, Japan, April 1998
33. Zhu, C.Z., Zhou, X., Satoh, S.: Bag-of-words against nearest-neighbor search for visual object retrieval. In: Proceedings of 2013 2nd IAPR Asian Conference on Pattern Recognition, pp. 626–630 (2013)