MARKOV NETWORK-BASED MULTIPLE CLASSIFIER FOR FACE IMAGE RETRIEVAL

Wonjun Hwang[+*], *Kyungshik Noh*[+], *Junmo Kim*[*]

[+]Samsung Advanced Institute of Technology, [*]KAIST, Korea
wj.hwang@samsung.com, kyung.roh@samsung.com, junmo@ee.kaist.ac.kr

ABSTRACT

We propose a new face-recognition framework to learn the relationship between multiple classifiers using a Markov network. For each image, we make three face models based on different distances between two eye locations. The novelty of the proposed method lies in that the method not only compares the query and target images at the three different levels, but also takes into account the statistical dependency between the three different models. This dependency is captured by a Markov network, which we describe by a graphical model, where query models are observation nodes, target models are hidden nodes, and the network line represents their relationships. For each observation–hidden node pair, we collect a set of target candidates that are most similar to the observation, and the relationship between the hidden nodes is captured in terms of the similarity between target images. Posterior probabilities at the three hidden nodes of the Markov network are computed by a belief-propagation algorithm. We evaluate the proposed method using FRGC ver 2.0, XM2VTS, BANCA, and PIE databases, which demonstrates its superiority under the untrained variations.

Index Terms**—** Markov Network, Multiple Face Model, Face Image Retrieval, Face Recognition

1. INTRODUCTION

Robust face recognition has many applications including biometrics, video surveillance, and human–robot interaction. Its performance [1][2][3][12][13] has gradually advanced through, in particular, the face recognition grand challenge (FRGC) competition [4]. This progress has largely been made by employing techniques that merge several classifiers using multiple feature sets of different characteristics, as in component-based methods, which extract features from separate spatial regions, and heterogeneous feature-based methods, which merge different domain features. Traditionally, classifiers for the different feature sets are trained independently, and the resulting similarity scores are summed up with the predefined weight factors. However, such a simple classifier-fusion scheme cannot reflect the correlation between the different models efficiently, and the predefined weights are easily over-fitted and sensitive to the variations in the training sets.

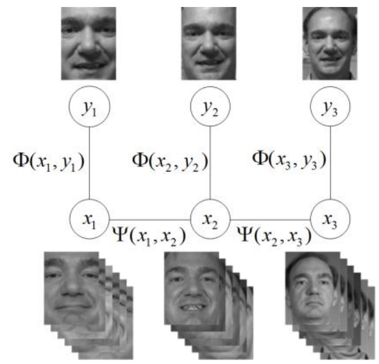

Fig. 1. Proposed Markov network for the multiple face models. From left to right, the interior, compensated, and exterior face models are shown. The circles represent network nodes and the lines mean statistical dependency between the nodes. In detail, **y** and **x** are observation and hidden nodes, respectively, and Φ and Ψ are compatibility functions that model the relationships between nodes.

Recently, Freeman et al. [5] proposed a learning-based method for low-level vision problems such as example-based super resolution. They generated a synthetic world of scenes and their corresponding rendered images, and modeled their relationship using a Markov network. Huang et al. [6] have proposed a hybrid face-recognition method that combines holistic and feature analysis-based approaches using a Markov random field (MRF) model. They divided a face image into small patches, and the relationship between image patches is captured by an MRF model. In the end, the binary decisions of patches are voted on for verification.

In this paper, we propose a Markov network-based face-recognition method to probabilistically model the relationship between not only a query face model and a target face model, but also between the target face models. First, given a query image, we make the multiple face models [3] according to different choices of eye distance during the normalization stage, and the boosted Gabor features [8] are extracted from each face model. With the Markov network, we unify the multiple face models by means of the network nodes and lines, as illustrated in Fig. 1. For each observation node, a query feature comes in as an input and then it retrieves the first n similar target candidates at the corresponding hidden node. We repeat this retrieval for each of the three query images separately. As a

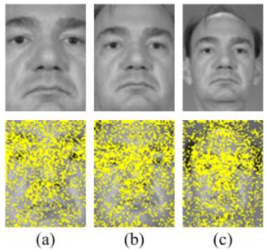

Fig. 2. Yellow dots show the spatial localities of the 2,400 features, $c = \{0.0\}$ and $\sigma = \{\pi\}$, selected by boosting theory in the (a) interior, (b) compensated, and (c) exterior face models.

result, the three query images, though they are identical images at different scales, have their own lists of retrieved target images that are not identical in general; thereby, they complement each other. Hidden nodes are connected by the network lines and we calculate the relationship between the hidden nodes. The posterior probability at each hidden node is easily computed by the belief-propagation [5] algorithm.

2. EXTENDED CURVATURE GABOR CLASSIFIER BUNCH

We simplify the extended curvature Gabor (ECG) classifier bunch [8] for the multiple face models. We only adopt four curvature parameters, $c \in \{0.0, 0.05, 0.1, 0.2\}$, and one radius parameter $\sigma = \pi$. For each choice of curvature parameter, we select 2,400 feature candidates by employing the boosting theory, and further reduce the dimension of the feature vector to 221 by linear discriminant analysis (LDA). We repeat this process for the four different curvature parameters, thereby constructing a feature vector of dimensionality $m = 221 \times 4$. In this paper, we have three face models [3], the interior, the compensated, and the exterior face models, with different eye distances within a regular face image region. Fig. 2 shows examples of the selected localities. Note that most features are found around the eyes and a nose, as described in [8].

3. MARKOV RANDOM FIELD (MRF)

Given a query image, $\mathbf{y}$, we would like to find the most similar target image $\mathbf{x}$ from the enrolled image set in face retrieval. This problem can be cast as a maximization of the posterior probability,

$$\begin{aligned} P(\mathbf{x}|\mathbf{y}) &= cP(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N) \\ &= c\prod_{(i,j)} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \prod_i \phi_i(\mathbf{x}_i, \mathbf{y}_i), \end{aligned} \quad (1)$$

where c is a normalization constant, ψ and ϕ are compatibility functions, and the product is over all neighboring pairs of N nodes. In this paper, we generate three face models, such as the multiple face model [3], and assign each face model to one node of the Markov network, where query face models are assigned to observation nodes, and target face models to hidden nodes. We use the belief-propagation algorithm to find the optimal solution, which iteratively updates message m_{ij} from node i to node j, and its equation is

$$\mathbf{m}_{ij}(\mathbf{x}_j) = \sum_{\mathbf{x}_i} \psi_{ij}(\mathbf{x}_i, \mathbf{x}_j) \prod_{k \neq j} \mathbf{m}_{ki}(\mathbf{x}_i) \phi_i(\mathbf{x}_i, \mathbf{y}_i). \quad (2)$$

The marginal probability b_i for a target $\mathbf{x}_i$ at a node i is derived by

$$b_i(\mathbf{x}_i) = \prod_k \mathbf{m}_{ki}(\mathbf{x}_i) \phi_i(\mathbf{x}_i, \mathbf{y}_i). \quad (3)$$

Ideally, the summation in (3) should be over all target candidates, but we limit the size of target candidates (denoted by n, whose value is empirically chosen to be 300), due to the computational complexity. The message and posterior probabilities are approximated by using the summation over the n most significant target candidates, which contribute most to the messages and probabilities.

4. PROPOSED MARKOV NETWORK-BASED RECOGNITION FRAMEWORK

We can use three face models that focus on learning complementary face models from the internal features to the external features. One simple way to handle them is to merge the score values of the different features, as in [3][7]. As a more advanced approach, in this paper, we model the spatial relationship from the internal to the external face models based on the Markov network. We do not divide an image into several patches as in [5][6] but, as shown in Fig. 1, we define the multiple face models of a query image as the interior $\mathbf{y}_1$, the compensated $\mathbf{y}_2$, and the exterior face model $\mathbf{y}_3$. Each observation node $\mathbf{y}_i$ has its own corresponding hidden node, $\mathbf{x}_i$. At each hidden node, we find a set of n target candidates that are most similar to the observation. The neighboring hidden nodes also have their own sets of target candidates. Now note that the hidden nodes are connected to each other in the graphical model. For example, the interior face model is related to the compensated face model, which, in turn, is also connected to the exterior face model. The relationship between hidden node $\mathbf{x}_i$ and $\mathbf{x}_j$ is evaluated for pairs of realizations, $(\mathbf{x}_{ip}, \mathbf{x}_{jq})$, $1 \le p \le n$, $1 \le q \le n$.

For a given query $\mathbf{y}_i$, the query–target compatibility function, $\Phi(\mathbf{x}_i, \mathbf{y}_i)$, is evaluated for n target candidates of $\mathbf{x}_i = \{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in}\}$, thus generating a vector in $\Re^{n \times 1}$. The query–target compatibility function is the traditional face-retrieval task. Using the normalized correlation method, the n most similar face images are retrieved from each node. However, the target–target compatibility function, $\Psi(\mathbf{x}_i, \mathbf{x}_j)$, is evaluated for $n \times n$ pairs of $(\mathbf{x}_i, \mathbf{x}_j)$, thus generating a matrix in $\Re^{n \times n}$. We define the compatibility function between the target nodes i and j as,

$$\psi_{ij}(x_i, x_j) = \exp\left(-\frac{|s_{ij}(x_i, x_j) - 1|^2}{2\sigma^2}\right), \quad (4)$$

where σ is a noise parameter. Now we need to define the similarity measure, $s_{ij}(\mathbf{x}_{ip}, \mathbf{x}_{jq})$, but $\mathbf{x}_{ip}$ and $\mathbf{x}_{jq}$ are from different face models and we cannot directly compare their corresponding features, $f^{x_{ip}}$ and $f^{x_{jq}}$. To address this

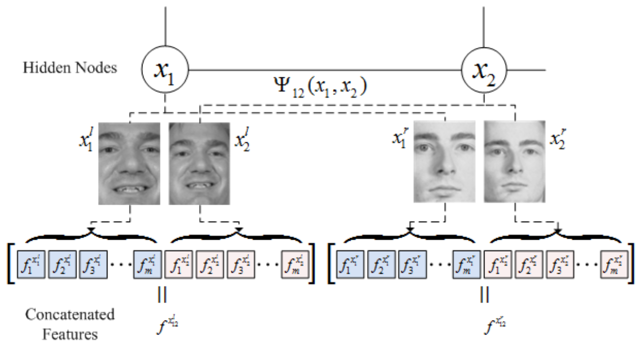

Fig. 3. The similarity between instances $\mathbf{x}_{1p}$ and $\mathbf{x}_{2q}$ of different hidden nodes $\mathbf{x}_1$and $\mathbf{x}_2$ is computed by comparing two concatenated features, $\bar{\mathrm{f}}^{\mathrm{x}_{1p}} = [\mathrm{f}^{\mathrm{x}_{1p}}\, \mathrm{f}^{\bar{\mathrm{x}}_p}]$ and $\bar{\mathrm{f}}^{\mathrm{x}_{2q}} = [\mathrm{f}^{\bar{\mathrm{x}}_q}\, \mathrm{f}^{\mathrm{x}_{2q}}]$, where $\mathbf{x}_{1p}$ and $\mathbf{x}_{2q}$ are from the retrieved target sets at the 1st and 2nd hidden nodes, respectively, and $\bar{\mathbf{x}}_p$ and $\bar{\mathbf{x}}_q$ are additional images for comparison between different model images. The additional image $\bar{\mathbf{x}}_p$, which follows the compensated image model (the model for node $\mathbf{x}_2$), comes from the same original image from which the target instance $\mathbf{x}_{1p}$ is obtained, and $\mathrm{f}_k^{\mathrm{x}_{1p}}$ is the kth element of the feature vector, $\mathrm{f}^{\mathrm{x}_{1p}}$. Similarly, $\bar{\mathbf{x}}_q$ is simply a resized version of $\mathbf{x}_{2q}$.

problem, we propose using the concatenated features, $\bar{\mathrm{f}}^{\mathrm{x}_{ip}}$ and $\bar{\mathrm{f}}^{\mathrm{x}_{jq}}$, as shown in Fig. 3. They are concatenated features that use the neighbor nodes, and we can compare them by using the normalized correlation,

$$\mathrm{s}_{ij}(\mathbf{x}_{ip}, \mathbf{x}_{jq}) = < \bar{\mathrm{f}}^{\mathrm{x}_{ip}}, \bar{\mathrm{f}}^{\mathrm{x}_{jq}} > / \|\bar{\mathrm{f}}^{\mathrm{x}_{ip}}\| \|\bar{\mathrm{f}}^{\mathrm{x}_{jq}}\|. \quad (5)$$

Belief propagation is a message-passing algorithm based on the graphical model. From equations (2) and (3), we derive the following equations for the MRF model of Fig. 1.

$$\mathrm{m}_{12}(\mathbf{x}_2) = \textstyle\sum_{\mathbf{x}_1} \psi_{12}(\mathbf{x}_1, \mathbf{x}_2)\phi_1(\mathbf{x}_1, \mathbf{y}_1), \quad (6)$$
$$\mathrm{m}_{23}(\mathbf{x}_3) = \textstyle\sum_{\mathbf{x}_2} \psi_{23}(\mathbf{x}_2, \mathbf{x}_3)\mathrm{m}_{12}(\mathbf{x}_2)\phi_2(\mathbf{x}_2, \mathbf{y}_2), \quad (7)$$
$$\mathrm{m}_{32}(\mathbf{x}_2) = \textstyle\sum_{\mathbf{x}_3} \psi_{32}(\mathbf{x}_3, \mathbf{x}_2)\phi_3(\mathbf{x}_3, \mathbf{y}_3), \quad (8)$$
$$\mathrm{m}_{21}(\mathbf{x}_1) = \textstyle\sum_{\mathbf{x}_2} \psi_{21}(\mathbf{x}_2, \mathbf{x}_1)\mathrm{m}_{32}(\mathbf{x}_2)\phi_2(\mathbf{x}_2, \mathbf{y}_2). \quad (9)$$

The marginal probabilities are

$$\mathrm{b}_1(\mathbf{x}_1) = \mathrm{m}_{21}(\mathbf{x}_1)\phi_1(\mathbf{x}_1, \mathbf{y}_1), \quad (10)$$
$$\mathrm{b}_2(\mathbf{x}_2) = \mathrm{m}_{12}(\mathbf{x}_2)\mathrm{m}_{32}(\mathbf{x}_2)\phi_2(\mathbf{x}_2, \mathbf{y}_2), \quad (11)$$
$$\mathrm{b}_3(\mathbf{x}_3) = \mathrm{m}_{23}(\mathbf{x}_3)\phi_3(\mathbf{x}_3, \mathbf{y}_3). \quad (12)$$

As the posterior probability $\mathrm{p}(\mathbf{x}_i | \mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ is proportional to $\mathrm{b}_i(\mathbf{x}_i)$ at hidden node $\mathbf{x}_i$, we can use these three marginal probabilities— $\mathrm{b}_1(\mathbf{x}_1)$, $\mathrm{b}_2(\mathbf{x}_2)$, and $\mathrm{b}_3(\mathbf{x}_3)$—for face recognition. We calculate the similarity value between the outputs of the two nodes, which then updates the result of the current node. Therefore, the similar candidates of both nodes are more weighted in the final result. In this respect, we can think of it as a kind of clustering-based face-recognition method.

5. EXPERIMENTAL RESULT AND DISCUSSION

To evaluate the performance of the proposed method, we use the FRGC ver 2.0 database [4]. In the FRGC, the training set consists of 12,776 images from 222 subjects. EXP1 is designed to measure performances of frontal face recognition under the controlled illumination condition with 16,028 query and target images, and EXP4 is a more practical protocol with 8,014 uncontrolled query images, and 16,028 controlled target images. Face-recognition accuracy is measured by the 1st rank of the cumulative match characteristic (CMC) curve. The performances are reported by three receiver-operating characteristic (ROC) curves obtained from three different subsets, which are different time lapses (ROC1, ROC2, and ROC3).

To generate the multiple face images, first we normalize an input image into 60×80 images with three different eye distances, for example, 24, 32, and 40 pixels. We can calculate three classifiers—Ω_1, Ω_2, and Ω_3—by using the well-known, traditional fusion method for Min, Max, Sum, and the likelihood rate [7]. Similarly, we have three marginal probabilities that are b_1, b_2, and b_3 from the proposed method. Fig. 4 shows the performances of the individual classifiers, and the best accuracies are achieved by the proposed method in both EXP1 and EXP4. In the

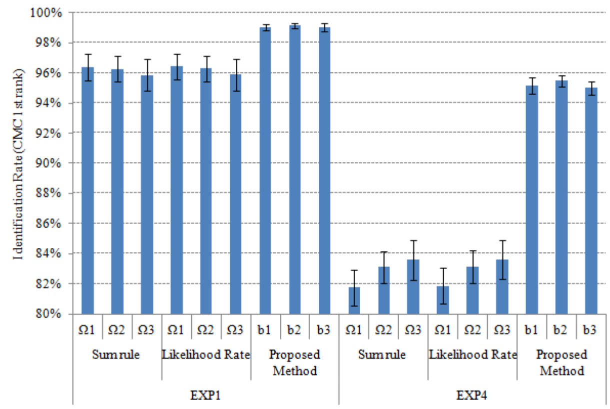

Fig. 4. The average identification rates of three face models are shown by the 1st rank values of the CMC curve. The bars are the standard deviations of the three ROCs.

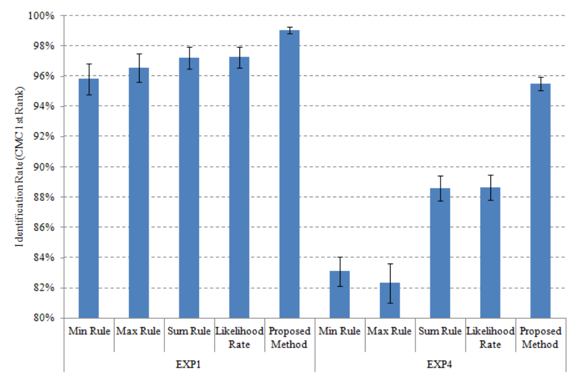

Fig. 5. Three face models are fused by Min, Max, Sum, likelihood rate, and the proposed method in EXP1 and EXP4 of the FRGC ver 2.0 database.

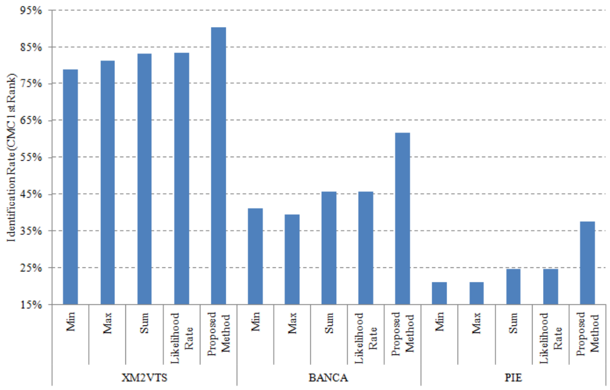

Fig. 6. Identification rates corresponding to Min, Max, Sum, the likelihood rate, and the proposed method are shown using the (a) XM2VTS, (b) BANCA, and (c) PIE databases.

sum rule and likelihood rate method, the exterior face model, Ω_3, achieves the best result, followed by Ω_2 and then Ω_1 in EXP4. However, this tendency is reversed in EXP1, and Ω_3 shows the worst result. We can therefore deduce that the eye, nose, and mouth components of the interior face model are useful in the controlled illumination setting, but under uncontrolled conditions, the exterior features (such as hairstyle) could be utilized for improving the recognition rate. Regarding the proposed method, compared with the other face models, b_1 and b_3, the compensated face model b_2 shows the best result. The main reason why b_2 of the proposed method shows good results, unlike Ω_2 of the other methods, is that the node $\mathbf{x}_2$ is connected to two neighboring nodes, whereas, $\mathbf{x}_1$ and $\mathbf{x}_3$ only have one neighboring node. Fig. 5 shows the performances of the merged classifiers. We can see that such a merging always leads to better accuracy than individual classifiers. In detail, average recognition rates of 83.10%, 82.32%, 88.60%, and 88.65% are reached by the Min, Max, Sum, and the likelihood rate-based methods, respectively, in EXP4. The proposed face models are simply merged by the sum rule, and this achieves an average recognition rate of 95.51% in EXP4.

We furthermore evaluate the accuracy changes of the proposed method according to the untrained variations. For this purpose, we use three face databases: XM2VTS [10], BANCA [9], and PIE [11]. First, we train the face models using only the FRGC ver 2.0 training set and do not add any other database as a training set. The XM2VTS, BANCA, and PIE databases are only used as test sets. We do not follow the evaluation protocols for these databases, in particular, for the XM2VTS and BANCA databases, but we do make use of all the images as test sets. In the PIE database, the query set is illumination and pose changes, and the target set is the frontal face with flash changes. Note that we could observe the performance degradations under different situations. For example, in the FRGC ver 2.0 training set, there are no variations with respect to facial-pose changing, but that the PIE database has plenty of pose variations. In Fig. 6, the proposed method always achieves 5–16% better accuracy than the other methods. The overall performances of all classifiers drop in comparison with FRGC EXP1 and EXP4 because of the untrained variations, such as the pose and illumination changes. Nevertheless, the proposed method always shows better results than the other methods, indicating that the superiority of the proposed method is relatively generalized.

6. CONCLUSION

We proposed a novel face-recognition method based on multiple face models and the corresponding Markov network. Three face models of a query image are assigned to the three observation nodes, and each hidden node corresponds to target images. The statistical dependency between the hidden nodes is calculated by measuring similarities between target images at neighboring hidden nodes via comparing the concatenated features. We retrieve the most similar target images from the database using a face model of a query image, and between a hidden–hidden node pair, we calculate the similarities between the retrieved target images. Hence, the resulting inference mechanism can be viewed as a kind of clustering-based face-recognition method. The proposed method achieves high accuracy in the FRGC ver 2.0, XM2VTS, BANCA, and PIE databases, compared with the other well-known face-recognition methods.

7. REFERENCES

[1] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical Ensemble of Global and Local classifiers for Face Recognition," IEEE Trans. on Image Processing, Vol. 18, No. 8, pp. 1885-1896, Aug. 2009.

[2] Z. Liu and C. Liu, "Robust Face Recognition Using Color Information," Advances in Biometrics, Vol. 5558/2009, pp. 122-131, 2009.

[3] W. Hwang, G. Park, J. Lee, and S. Kee, "Multiple Face Model of Hybrid Fourier Feature for Large Face Image Set," IEEE Conf. on Computer Vision and Pattern Recognition, New York, USA, Jun., 2006.

[4] P. J. Phillips, P. J. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the Face Recognition Grand Challenge," Computer Vision and Pattern Recognition, Vol. 1, pp. 947-954, Jun. 2005.

[5] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning Low-level Vision," Inter. Journal of Computer Vision, Vol. 40, No. 1, pp. 25-47, Jul. 2000.

[6] R. Huang, V. Pavlovic, and D. N. Metaxas, "A Hybrid Face Recognition Method Using Markov Random Fields," Inter. Conf. on Pattern Recognition, Vol. 3, pp. 157-160, Aug. 2004.

[7] S. Prabhakar and A. K. Jain, "Decision-level Fusion in Figure-print Verification," Pattern Recognition, Vol. 4, No. 35, pp. 861-874, Apr. 2002.

[8] W. Hwang, X. Huang, K. Noh, and J. Kim, "Face Recognition System Using Extended Curvature Gabor Classifier Bunch for Low-resolution Face Image," IEEE CVPR Workshop, pp. 17-24, Jun. 2011.

[9] E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, and et al, "The BANCA Database and Evaluation Protocol," Audio- and

Video-based Biometric Person Authentication, Vol. 2688/2003, 2003.
[10] K. Messer, J. Kittler, M. Sadeghi, S. Marcel, C. Marcel, and et al, “Face Verification Competition on the XM2VTS Database,” Audio- and Video-based Biometric Person Authentication, Vol. 2688/2003, 2003.
[11] T. Sim, “The CMU Pose, Illumination, and Expression (PIE) Database,” IEEE Inter. Conf. on Automatic Face and Gesture Recognition, pp. 46-51, May 2002.
[12] P. Parveen and B. Thuraisingham, “Face Recognition Using Multiple Classifiers,” IEEE International Conference on Tools with Artificial Intelligence, pp. 179-186 , Nov. 2006.
[13] D. Skocaj, A. Leonardis, H. Bishof, “Weighted and robust learning of subspace representations,” Pattern Recognition, Vol. 40, No. 5, pp. 1,556-1,569, May, 2007.