

Multi-Frame Face Classification with Decision-Level Fusion based on Photon-Counting Linear Discriminant Analysis

Seokwon Yeom

Division of Computer and Communication Engineering, Daegu University, Gyeongsan, Korea

The logo for the International Journal of Fuzzy Logic and Intelligent Systems (IJFIS) is displayed in a large, bold, blue font.

Abstract

Face classification has wide applications in security and surveillance. However, this technique presents various challenges caused by pose, illumination, and expression changes. Face recognition with long-distance images involves additional challenges, owing to focusing problems and motion blurring. Multiple frames under varying spatial or temporal settings can acquire additional information, which can be used to achieve improved classification performance. This study investigates the effectiveness of multi-frame decision-level fusion with photon-counting linear discriminant analysis. Multiple frames generate multiple scores for each class. The fusion process comprises three stages: score normalization, score validation, and score combination. Candidate scores are selected during the score validation process, after the scores are normalized. The score validation process removes bad scores that can degrade the final output. The selected candidate scores are combined using one of the following fusion rules: maximum, averaging, and majority voting. Degraded facial images are employed to demonstrate the robustness of multi-frame decision-level fusion in harsh environments. Out-of-focus and motion blurring point-spread functions are applied to the test images, to simulate long-distance acquisition. Experimental results with three facial data sets indicate the efficiency of the proposed decision-level fusion scheme.

Keywords: Decision-level fusion, Data fusion, Face classification, Face recognition, Photon-counting, Linear discriminant analysis

1. Introduction

Face classification has many applications in security monitoring and intelligent surveillance, as well as robot vision, image and video retrieval, and human-machine interfaces [1-3]. However, it is challenging to classify a facial image acquired in an uncontrolled setting, such as those captured at long distances. Unexpected blurring and noise may occur, in addition to conventional distortions caused by pose, illumination, and expression changes. To address these issues, various classifiers have been developed based on statistical analysis, including Fisher linear discriminant analysis (LDA) combined with principal component analysis (PCA) [4], (often referred to as “Fisherfaces”), as well as the “Eigenfaces” method, which uses only PCA [1]. Typically, the number of training images is much less than the number of pixels.

Received: Nov. 4, 2014
Revised : Nov. 19, 2014
Accepted: Dec. 5, 2014

Correspondence to: Seokwon Yeom
(yeom@daegu.ac.kr)
©The Korean Institute of Intelligent Systems

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Thus, the Fisher LDA requires a dimensionality reduction such as PCA in order to avoid the singularity problem, often referred to as the “small sample size problem.” However, photon-counting (PC) LDA does not suffer the singularity problem associated with a small sample size [5]. Originally, PC-LDA had been developed to train grayscale images and classify a photon-limited image obtained under low illumination. However, it has been shown that PC-LDA is also suitable for classifying grayscale images, which can be obtained by a visible camera [6].

Decision-level fusion is a high-level data fusion technique [7, 8]. It aims to increase classification accuracy by combining multiple outputs from multiple data sets. Compared to single frames, multi-frames contain additional information acquired from varying spatial or temporal settings, as illustrated in Figure 1. Various fusion rules such as maximum, averaging, and majority-voting rules have been studied in the literature [9, 10]. Bayesian estimation and Dempster-Shafer evidential reasoning are often adopted for decision-level fusion [11]. In [12], preliminary results are provided for multi-frame recognition with several data sets.

In this paper, multi-frame decision-level fusion with PC-LDA is discussed. Decision-level fusion involves three stages: score normalization, score validation, and score combination. After the scores are normalized, candidate scores are selected using a screening process (score validation). Subsequently, the scores representing the classes are combined to render a final decision using a fusion rule (score combination). The validation stage screens out “bad” scores that can degrade classification performance. The maximum, averaging, and majority voting fusion rules are investigated in the experiments. Three facial image datasets (ORL, AR, Yale) [13-15] are employed to verify the effectiveness of the proposed decision-level fusion scheme.

The remainder of the paper is organized as follows. PC-LDA is discussed in Section 2. Section 3 describes decision-level fusion. The experimental results are presented in Section 4. The conclusion follows in Section 5.

2. Photon Counting LDA

This section briefly describes PC-LDA. PC-LDA realizes the Fisher criterion using the Poisson distribution, which characterizes the semi-classical photo-detection model [16]. A PC vector \mathbf{y} is a random feature vector corresponding to a normalized image vector \mathbf{x} . Thus, the dimensions of \mathbf{x} and \mathbf{y} are the same value, which is the number of pixels d ; y_i is the i -th

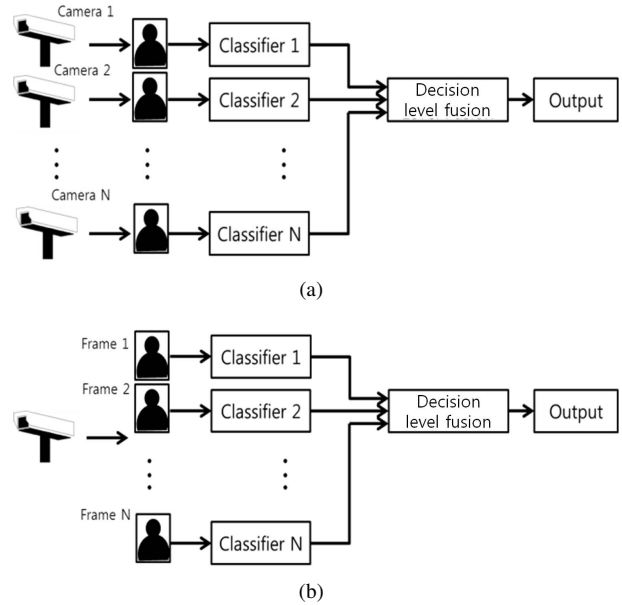


Figure 1. Configurations of varying (a) spatial setting, (b) temporal setting.

component of \mathbf{y} , and it follows the independent Poisson distribution with the parameter $N_p x_i$, that is, $y_i \sim \text{Poisson}(N_p x_i)$. It is noted that x_i is the normalized intensity at a pixel i such that $\sum_{i=1}^d x_i = 1$, and N_p indicates the total number of average photo-counts because the following equation is valid:

$$\sum_{i=1}^d E(y_i) = \sum_{i=1}^d N_p x_i = N_p.$$

The between-class covariance measures the separation of classes as

$$\begin{aligned} \Sigma_{yy}^B &= E_j \left[(\boldsymbol{\mu}_{y|j} - \boldsymbol{\mu}_y)(\boldsymbol{\mu}_{y|j} - \boldsymbol{\mu}_y)^t \right] \\ &= N_p^2 \Sigma_{xx}^B, \end{aligned} \tag{1}$$

where the class-conditional mean and the mean vectors are derived as $\boldsymbol{\mu}_{y|j} = N_p \boldsymbol{\mu}_{x|j}$ and $\boldsymbol{\mu}_y = N_p \boldsymbol{\mu}_x$, respectively; j indicates a class, and superscript t denotes a matrix transpose. The within-class covariance matrix measures the concentration of members in the same class as

$$\begin{aligned} \Sigma_{yy}^W &= E_j \left\{ E_{y|j} \left[(\mathbf{y} - \boldsymbol{\mu}_{y|j})(\mathbf{y} - \boldsymbol{\mu}_{y|j})^t | j \right] \right\} \\ &= N_p \text{diag}(\boldsymbol{\mu}_x) + N_p^2 \Sigma_{xx}^W, \end{aligned} \tag{2}$$

where $\text{diag}(\cdot)$ denotes a diagonal matrix. Thus, the following

Fisher criterion can be derived:

$$\begin{aligned}
 W_P &= \arg \max_{W \in R^{d \times r}} \frac{|W^t \Sigma_{yy}^B W|}{|W^t \Sigma_{yy}^W W|} \\
 &= \arg \max_{W \in R^{d \times r}} \frac{|W^t \Sigma_{xx}^B W|}{|W^t [\text{diag}(\boldsymbol{\mu}_x)/N_p + \Sigma_{xx}^W] W|},
 \end{aligned} \tag{3}$$

where the column vectors of W_P are equivalent to the eigenvectors of $[\text{diag}(\boldsymbol{\mu}_x)/N_p + \Sigma_{xx}^W]^{-1} \Sigma_{xx}^B$ corresponding to the non-zero eigenvalues. It is noted that $[\text{diag}(\boldsymbol{\mu}_x)/N_p + \Sigma_{xx}^W]$ is non-singular because of the non-zero components of $\boldsymbol{\mu}_x$.

The class decision can be made by maximizing a score function, as follows:

$$\hat{j} = \arg \max_{j=1, \dots, C} \text{score}(j), \tag{4}$$

where C is the number of classes. The normalized correlation is adopted as a score function:

$$\text{score}_y(j) = \frac{(W_p^t \mathbf{y}_u)^t (W_p^t \boldsymbol{\mu}_{y|j})}{\|W_p^t \mathbf{y}_u\| \|W_p^t \boldsymbol{\mu}_{y|j}\|}. \tag{5}$$

The photo-counting vector \mathbf{y}_u of an unlabeled object is required for class decisions, as depicted in Eq. (5). Alternatively, \mathbf{y}_u can be estimated with the intensity image vector \mathbf{x}_u . Because the minimum mean-squared error (MMSE) estimation is the conditional mean [17], a point estimation of y_{ui} becomes $E(y_{ui}|x_{ui}) = N_p x_{ui}$, where y_{ui} and x_{ui} are the i -th component of \mathbf{y}_u and \mathbf{x}_u , respectively. Thus, Eq. (5) is equivalent to the following score function:

$$\text{score}_x(j) = \frac{(W_p^t \mathbf{x}_u)^t (W_p^t \boldsymbol{\mu}_{x|j})}{\|W_p^t \mathbf{x}_u\| \|W_p^t \boldsymbol{\mu}_{x|j}\|}. \tag{6}$$

The mean-squared (MS) error is the same as the variance of y_{ui} , which is $N_p x_{ui}$. The MS error increases as N_p increases; however, the PC-LDA converges to the Fisher LDA as N_p goes to the infinite as

$$\lim_{N_p \rightarrow \infty} \frac{|W^t \Sigma_{xx}^B W|}{|W^t [\text{diag}(\boldsymbol{\mu}_x)/N_p + \Sigma_{xx}^W] W|} = \frac{|W^t \Sigma_{xx}^B W|}{|W^t \Sigma_{xx}^W W|}. \tag{7}$$

Two performance measures are calculated to evaluate the performance of the classifiers. One is the probability of correct decisions (P_D), and the other is the probability of false alarms (P_{FA}) [6]:

$$P_D(j) = \frac{\text{Number of decision for class } j}{\text{Number of test images in class } j}, \tag{8}$$

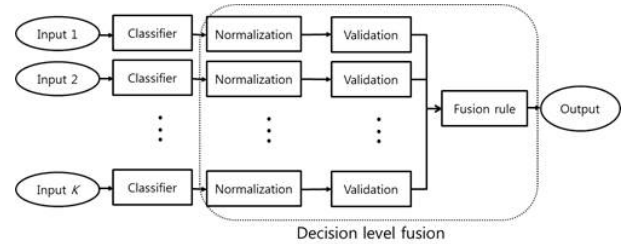


Figure 2. Block diagram showing decision-level fusion.

$$P_{FA}(j) = \frac{\text{Number of decision for class } j, \text{ but are not in class } j}{\text{Number of test images in all classes except for class } j}. \tag{9}$$

3. Decision-Level Fusion

Decision-level fusion is composed of three stages: score normalization, validation, and fusion rule processes; these are illustrated in Figure 2. The scores must be normalized if they are presented in different metric forms. The candidate scores are selected during the validation process. Finally, they are combined to create a new score, using a fusion rule. For the score validation, a score set S_k is composed of n_k scores selected from the output scores of a frame k as follows:

$$S_k = \{\text{score}_x(j_{(1)}; k), \dots, \text{score}_x(j_{(n_k)}; k)\}, \quad k = 1, \dots, K, \tag{10}$$

$$\begin{aligned}
 \text{score}_x(j_{(1)}; k) &> \dots > \text{score}_x(j_{(n_k)}; k) \\
 &\dots > \text{score}_x(j_{(C)}; k),
 \end{aligned} \tag{11}$$

where K is the total number of frames. S_1, \dots, S_K score sets are then reassigned to new sets $\tilde{S}_1, \dots, \tilde{S}_C$ as follows:

$$\tilde{S}_j = \{\text{score}_x(j; 1), \text{score}_x(j; 2), \dots, \text{score}_x(j; \tilde{n}_j)\} \quad j = 1, \dots, C, \tag{12}$$

where \tilde{n}_j is the number of scores for class j from all K frames. Therefore, $\bigcup_{k=1}^K S_k = \bigcup_{j=1}^C \tilde{S}_j$ and $\sum_{k=1}^K n_k = \sum_{j=1}^C \tilde{n}_j$ are held between the sets S_k and \tilde{S}_j . The following three fusion rules are adopted to compute the final score for class j :

$$\text{score}_{max}(j) = \max\{\tilde{S}_j\}, \tag{13}$$

$$\text{score}_{avg}(j) = \text{average}\{\tilde{S}_j\}, \tag{14}$$

$$\text{score}_{voting}(j) = \tilde{n}_j, \tag{15}$$

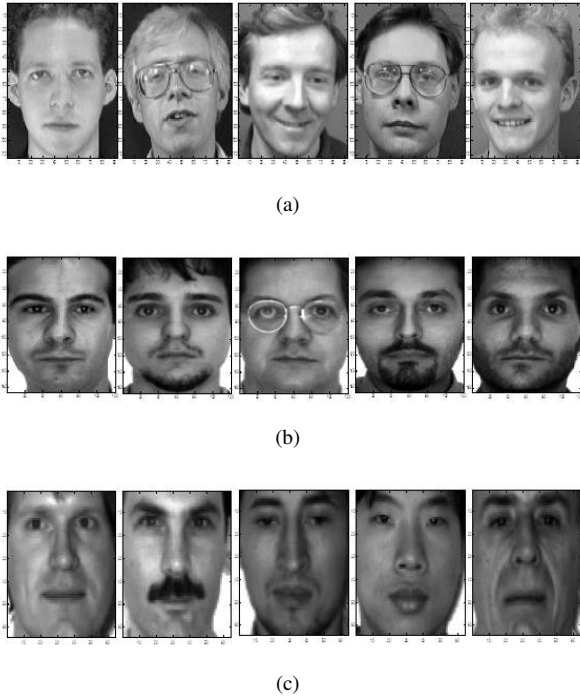


Figure 3. Sample images from (a) ORL, (b) AR, (c) Yale.

where Eqs. (13)-(15) represent maximum, averaging, and majority voting rules, respectively.

4. Experimental and Simulation Results

This section describes two types of experiments. The first involves the verification of PC-LDA with a single frame. In the second experiment, decision-level fusion is tested with artificially degraded test images.

4.1 Face Classification

Three facial image datasets were used for the performance evaluation: ORL [13], AR, [14], and Yale [1]. The MATLAB format was utilized for the Yale database [15]. Figure 3 shows the sample images of five classes from three datasets. The datasets contain 40, 100, and 15 classes, respectively; these classes respectively contain 10, 26, and 11 images. The dataset image sizes are 92×112 , 120×165 , and 64×64 pixels, respectively. Each database was divided into three validation sets, as shown in Table 1. For the single-frame experiment, each validation set was trained and all other validation sets were tested. For example, when three images (image indexes 1-3) in set V_1 of the ORL dataset were trained, the other seven

Table 1. Image index in validation sets

Validation set	ORL	AR	Yale
V_1	1-3	1-8	1-3
V_2	4-6	9-17	4-7
V_3	7-10	17-26	8-11

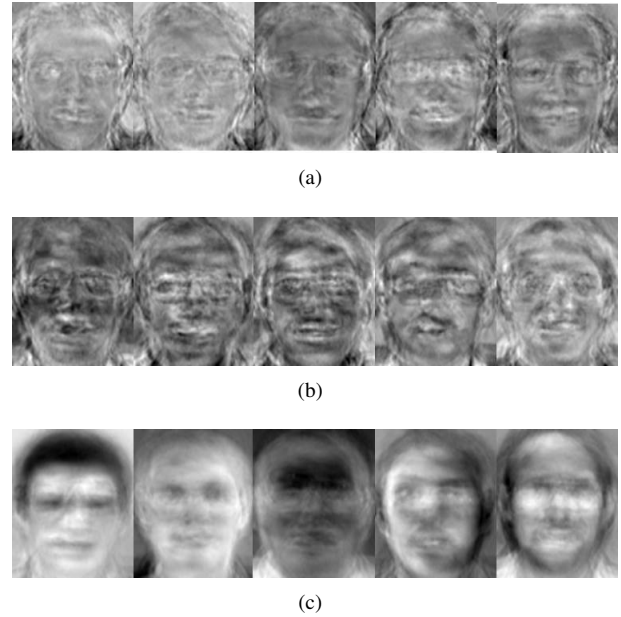


Figure 4. (a) Photon-counting linear discriminant analysis face, (b) Fisherface, (c) Eigenface.

images (image indexes 4-10) were tested. Figures 4 represents the five column vectors of the PC-LDA face, Fisherface, and Eigenface projection matrices, respectively, in the image scale; three images from set V_1 of the ORL dataset were trained to produce these results. As illustrated in the figures, the PC-LDA face presents the optimal structural diversity among the three classifiers, although the Eigen face method is more dependent on the intensity distribution, compared to the other methods. Figure 5 shows the average probability of detection (P_D) and average probability of a false alarm (P_{FA}) when each validation set is trained and other images are tested as a single frame. The results are compared with the Fisherface and Eigenface methods.

4.2 Decision-Level Fusion

For the decision-level fusion experiment, test images were blurred by out-of-focus and motion blurring point-spread functions, to simulate long-distance acquisitions. Out-of-focus im-

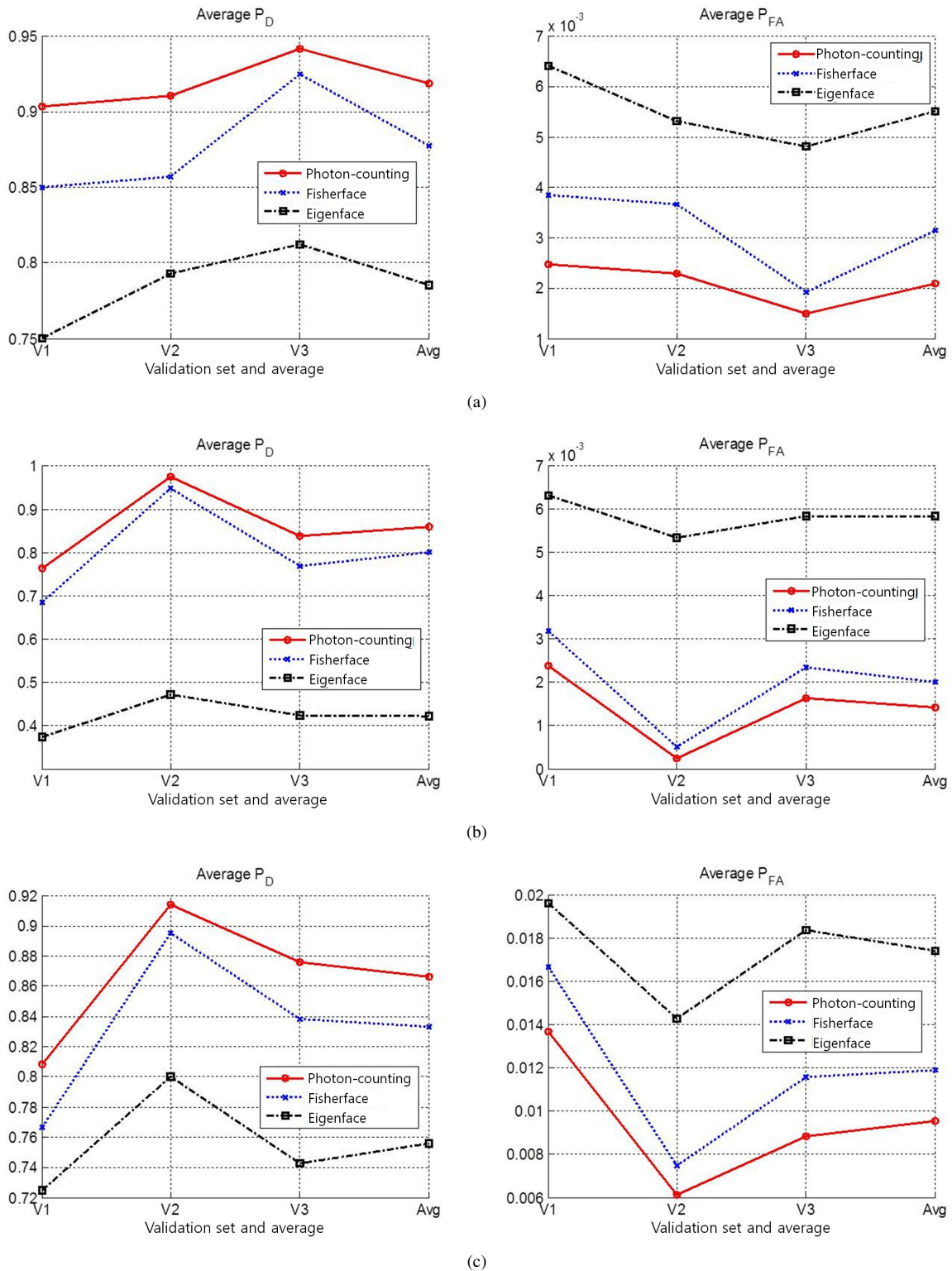


Figure 5. Single frame results of P_D and P_{FA} : (a) ORL, (b) AR, (c) YALE.

ages were rendered by applying circular averaging with an 8 pixel radius. Heavy motion blurring was rendered by a filter

approximating the linear motion of a camera for a distance of 20 pixels, with an angle of 45° in a counter-clockwise direction

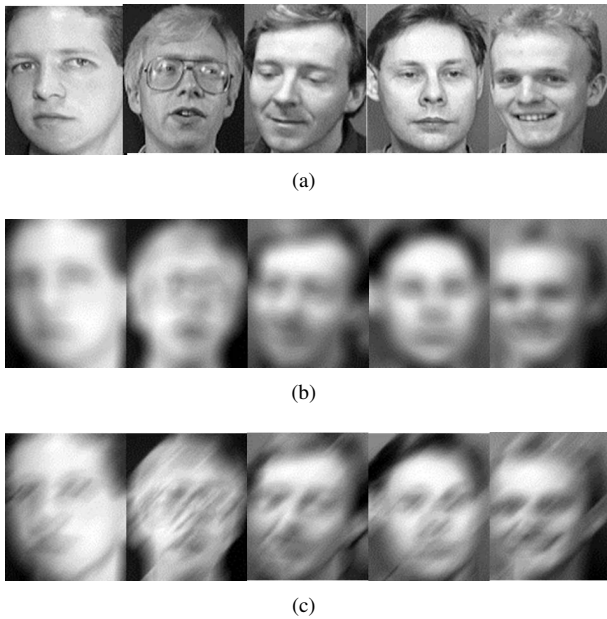


Figure 6. Sample test images from ORL: (a) original, (b) out-of-focus blurring, (c) motion blurring.

[6]. Figure 6 shows the sample test images from ORL after blur rendering.

It was assumed that one pair of test images in the validation set was obtained by multiple sensors; thus, the total number of frames (K) was set to two. For example, if the number of test images was seven in the single-frame experiment, the number of test pairs for the multi-frame fusion was 21 ($= {}_7C_2$). Figure 7 shows the average P_D and P_{FA} for the ORL, AR, and YALE datasets. The maximum rule produced the optimal results for the original images; however, the majority rule produced the optimal results when the images were degraded with the blurring functions.

5. Conclusions

This study investigated the effectiveness of a decision-level fusion system with multi-frame facial images. Three decision-level fusion schemes were investigated, following the score normalization and validation processes. Two types of blurring point-spread functions were applied to the test images, in order to simulate harsh conditions. The results indicated that the proposed data fusion scheme improved the classification performance significantly.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012R1A1A2008545).

References

- [1] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, Jul. 1997. <http://dx.doi.org/10.1109/34.598228>
- [2] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *Journal of the Optical Society of America A*, vol. 14, no. 8, pp. 1724-1733, Aug. 1997. <http://dx.doi.org/10.1364/JOSAA.14.001724>
- [3] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, Jan. 2004. <http://dx.doi.org/10.1109/TCSVT.2003.818349>
- [4] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed., New York, NY: Wiley, 2001.
- [5] S. Yeom, B. Javidi, and E. Watson, "Three-dimensional distortion-tolerant object recognition using photon-counting integral imaging," *Optics Express*, vol. 15, no. 4, pp. 1513-1533, Feb. 2007. <http://dx.doi.org/10.1364/OE.15.001513>
- [6] S. Yeom, "Photon-counting linear discriminant analysis for face recognition at a distance," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 12, no. 3, pp. 250-255, Sep. 2012. <http://dx.doi.org/10.5391/IJFIS.2012.12.3.250>
- [7] L. O. Jimenez, A. Morales-Morell, and A. Creus, "Classification of hyperdimensional data based on feature and decision fusion approaches using projection pursuit, majority voting, and neural networks," *IEEE Transactions on*

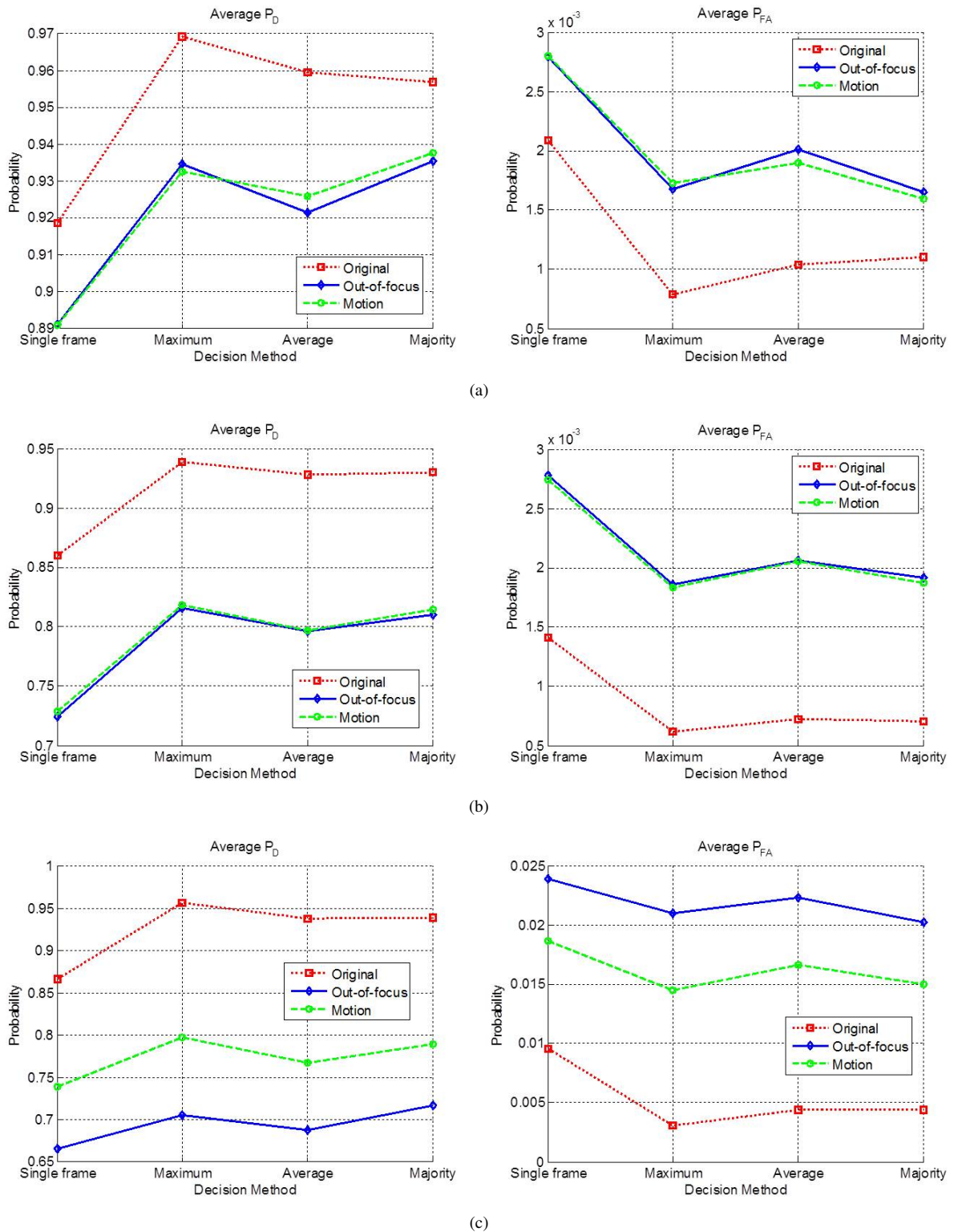


Figure 7. Decision-level fusion results of P_D and P_{FA} : (a) ORL, (b) AR, (c) YALE.

Geoscience and Remote Sensing, vol. 37, no. 3, pp. 1360-1366, May 1999. <http://dx.doi.org/10.1109/36.763300>

[8] S. Canavan, B. Johnson, M. Reale, Y. Zhang, L. Yin, and J. Sullins, "Evaluation of multi-frame fusion based

- face classification under shadow,” in *Proceedings of the 20th International Conference on Pattern Recognition*, Istanbul, Turkey, August 23-26, 2010, pp. 1265-1268. <http://dx.doi.org/10.1109/ICPR.2010.315>
- [9] M. Sadeghi, M. Samiei, and J. Kittler, “Fusion of PCA-based and LDA-based similarity measures for face verification,” *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. 1, pp. 647597, 2010. <http://dx.doi.org/10.1155/2010/647597>
- [10] P. Yanwei, Y. Nenghai, Z. Rong, R. Jiawei, and L. Zhengkai, “Fusion of SVD and LDA for face recognition,” in *Proceedings of the International Conference on Image Processing*, Singapore, October 24-27, 2004, pp. 1417-1420. <http://dx.doi.org/10.1109/ICIP.2004.1419768>
- [11] D. D. Freedman, “Overview of decision level fusion techniques for identification and their application,” in *Proceedings of the American Control Conference*, Baltimore, MD, June 29-July 1, 1994, pp. 1299-1303. <http://dx.doi.org/10.1109/ACC.1994.752269>
- [12] S. Yeom, “Decision-level fusion approach to face recognition with multiple cameras,” *SPIE Proceedings*, vol. 9120, pp. 91200G, May 2014. <http://dx.doi.org/10.1117/12.2053638>
- [13] F. S. Samaria and A. C. Harter, “Parameterisation of a stochastic model for human face identification,” in *Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, December 5-7, 1994, pp. 138-142. <http://dx.doi.org/10.1109/ACV.1994.341300>
- [14] A. M. Martinez and A. C. Kak, “PCA versus LDA,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228-233, Feb. 2001. <http://dx.doi.org/10.1109/34.908974>
- [15] A. S. Georghiadis, P. N. Belhumeur, and D. Kriegman, “From few to many: illumination cone models for face recognition under variable lighting and pose,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643-660, Jun. 2001. <http://dx.doi.org/10.1109/34.927464>
- [16] J. W. Goodman, *Statistical Optics*, 2nd ed., New York, NY: Wiley, 1985.
- [17] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York, NY: McGraw-Hill, Inc., 1991.



Seokwon Yeom received the M.S. and B.S. degrees in Electronics Engineering from Korea University, Seoul, South Korea and the Ph.D. degree in Electrical and Computer Engineering from the University of Connecticut in 2006. He is currently an associate professor in the Division of Computer and Communication Engineering at Daegu University in South Korea. His research interests are signal and image processing, optical information processing, and pattern recognition. He is now performing several research projects granted by Korea government and industries. Research Area: Signal and Image Processing, Optical Information Processing, Pattern Recognition. E-mail : yeom@daegu.ac.kr