

DETECTION AND RECOVERY OF OCCLUDED FACE IMAGES BASED ON CORRELATION BETWEEN PIXELS

Ji-eun Lee and Nojun Kwak

*Department of Electrical & Computer Engineering, Ajou University, Suwon, 443-749, South Korea
{mokona85, nojunk}@ajou.ac.kr*

Keywords: Correlation coefficient: Occluded face: Occlusion detection: Face image recovery.

Abstract: In this paper, we propose a method to detect and recover the occluded parts of face images using the correlation between pairs of pixels. In the training stage, correlation coefficients between every pairs of pixels are calculated using the occlusion-free training face images. Once a new face image is shown, the occluded area is detected and recovered using correlation coefficients obtained in the training stage. We compare the performance of the proposed method with the conventional method based on PCA. The results show that the proposed method detects and recovers occluded area with much smaller noises than the conventional PCA based method.

1 INTRODUCTION

Recently, a lot of surveillance cameras are used to obtain the identities of criminals. However, most of the criminals wear sunglasses or a mask, so it is very difficult to get the identity of the criminal using the currently available face recognition techniques. Therefore, the problem of detecting and recovering the occluded parts of a face image is a very important task to solve.

There has been works on detecting and recovering the occlusion of a face and the representative approach is to apply PCA (principal component analysis). These methods includes automatic eyeglasses removal from face images (Wu, 2004), reconstruction of the occluded parts by fast recursive PCA (Wang, 2007), application of probabilistic PCA (XiaoFeng, 2010), image completion (Efros, 1999; Kwatra, 2003; Jia, 2003; Sun, 2005; Komodakis, 2006) and so on (Hwang, 2003; Lin, 2007).

The outline of the conventional PCA based methods are as follows. Firstly, we reconstruct an occluded face image using a weight matrix composed of the eigenvectors of the training dataset consisting of non-occluded face images. Next, if the difference of pixel values between the reconstructed face image and the original occluded face image exceeds a threshold, that pixel is regarded as

occluded one. After detecting the occluded parts of a face image, the pixel values of the occluded parts are replaced by the pixel values of the reconstructed face image. Then, the above procedure is repeated on the output image of the previous step until the difference of pixel values between the reconstructed face image and the occluded face image is less than a threshold.

On the other hand, in this paper, instead of using PCA, we propose to use correlation between pixels in detecting and recovering the occluded parts of a face image. The proposed algorithm can be outlined as follows. Firstly, the proposed method calculates the correlation coefficients between every pairs of pixels using the training images. Secondly, we estimate the pixel values of occluded parts of each test images using the correlation coefficients. Suppose we do not know one pixel value but have the knowledge of the other pixel values. In this case, the unknown pixel value can be estimated by the other pixel values which are highly correlated with the pixel in question. If we apply this method to every pixels in the test image, we can obtain a reconstructed image whose pixel values are estimated from the other pixels. After that, if the difference of pixel values between the reconstructed face image and the occluded face image is over a threshold, those parts are considered as occluded ones and these pixels can be replaced by the ones

from the reconstructed image. This procedure can be applied repeatedly until the difference between the reconstructed and the occluded images are small.

As outlined above, this paper solves the problem of detecting and recovering the occluded face images using the correlation based method. The paper is organized as follows. In Section 2, we will review the conventional occlusion detection and recovery method based on PCA. In Section 3, the method based on correlation between pairs of pixels is proposed. Then we will compare the result of the proposed method with that of the conventional PCA based method through experiments in Section 4. Finally, discussions and conclusions follow in Section 5.

2 CONVENTIONAL PCA BASED METHOD

2.1 Training (PCA)

The purpose of training is to get a weight matrix consisting of enough number of the eigenvectors to reconstruct a face image.

The training set consisting of N labelled faces with d pixels can be represented as $X_{tr} = [x_1, x_2, \dots, x_N] \in \mathbb{R}^{d \times N}$ where each image x_n is represented as a column vector.

The next step is to get the mean image of the entire face images and subtracting the mean image from each face image.

$$\bar{x} = \frac{1}{N} \sum_{n=1}^N x_n \quad (1)$$

$$\Phi_n = x_n - \bar{x}, \quad n = 1, 2, \dots, N \quad (2)$$

Then we get a covariance matrix as the following:

$$C = \frac{1}{N} \sum_{n=1}^N \Phi_n \Phi_n^T \quad (3)$$

The d' largest eigenvalues $\lambda_1, \dots, \lambda_{d'}$ are selected by eigen-analysis and a weight matrix $W \in \mathbb{R}^{d \times d'}$ is obtained from the eigenvectors corresponding to the eigenvalues as follows:

$$W = [w_1, w_2, \dots, w_{d'}] \quad (4)$$

Here, w_n is n^{th} eigenvector corresponding to the eigenvalue λ_n .

2.2 Detection and Recovery of Occluded Face Images

Reconstruction: Each test face image x is projected to a d' -dimensional subspace and then a reconstructed image x' is obtained by using this projection as follows:

$$u_i = w_i^T (x - \bar{x}), \quad i = 1, 2, \dots, d' \quad (5)$$

$$\begin{aligned} x' &= \bar{x} + w_1 u_1 + \dots + w_{d'} u_{d'} \\ &= \bar{x} + \sum_{i=1}^{d'} w_i u_i \end{aligned} \quad (6)$$

Occlusion detection: The reconstructed face image x' and the original test face image x are compared to decide whether a pixel is occluded or not. This is achieved by checking the difference between pixel values of x and x' at the same pixel is more than a threshold ϵ or not. This procedure is described in the following equation:

$$\begin{aligned} |x'_i - x_i| &\leq \epsilon \implies i \in O^c \\ &> \epsilon \implies i \in O. \end{aligned} \quad (7)$$

Here, i denotes the specific location of the pixel and x_i and x'_i are i -th pixel values of x and x' respectively. The set O is the set of occluded pixels and O^c is the complement of O which corresponds to the set of non-occluded pixels.

Recovery: After detecting occluded parts from the face image, occluded parts are filled with the reconstructed image x' while non-occluded parts remains the same as the original image x . The resulting recovered image x'' is as follows:

$$x'' = \begin{cases} x_i & \text{if } i \in O^c \\ x'_i & \text{if } i \in O \end{cases} \quad (8)$$

After the recovery stage, x is replaced with x'' and the process is repeated until the difference between x and x'' becomes small enough. The final recovered image is obtained at the end of iteration when the process converges.

3 THE PROPOSED METHOD

3.1 Correlation Coefficient

The basic idea of the proposed method is that the pixel values of the occluded part can be predicted

from the pixel values of the non-occluded part which are highly related to the ones in question.

Suppose each pixel value be a random variable. Then the relationship between a pair of pixels can be measured by the correlation coefficient.

The correlation coefficient ρ_{ij} between two pixels x_i and x_j is defined as

$$\rho_{ij} = \frac{c_{ij}}{\sigma_i \sigma_j} = \frac{E[(x_i - \mu_i)(x_j - \mu_j)]}{\sqrt{E[(x_i - \mu_i)^2]} \sqrt{E[(x_j - \mu_j)^2]}} \quad (9)$$

Here, C_{ij} is the covariance between two random variables and σ_i, σ_j are the standard deviations of each random variable. The μ_i and μ_j are the mean values of x_i and x_j respectively and $E[\cdot]$ is the expectation operation. As in (9), the correlation coefficient is defined as a covariance between two random variables divided by the product of standard deviations of each variable, whose value should be from -1 to 1. If the correlation coefficient is close to 1 or -1, those two random variables are much related to each other and if it is close to 0, those two are hardly related at all.

Because each pixel can be considered as a random variable in the image, one image can be considered as a d-dimensional random vector and the correlation of each pixel can be obtained by formula (9).

Normally, the expectation of random variables cannot be calculated without knowing the underlying joint distribution. However, it can be estimated by the samples and the estimated value of the correlation coefficient (9) using N samples can be calculated as follows.

$$\hat{\rho}_{ij} = \frac{\sum_{n=1}^N (x_{in} - \bar{x}_i)(x_{jn} - \bar{x}_j)}{\sqrt{\sum_{n=1}^N (x_{in} - \bar{x}_i)^2} \sqrt{\sum_{n=1}^N (x_{jn} - \bar{x}_j)^2}} \quad (10)$$

Here, x_{in} is the i^{th} pixel value of the n^{th} training image and \bar{x}_i and \bar{x}_j are i^{th} and j^{th} sample mean of pixel values respectively.

Figure 1 shows some correlation coefficient maps of a face. The pixel marked X is the reference pixel (cheek, eye, nose, and mouse) from which the correlation is calculated with all the other pixels.

In the figure, white area corresponds to pixels that have high absolute value of the correlation coefficient with the reference pixel, while black area corresponds to pixels that have low absolute value of the correlation coefficient with the reference pixel.

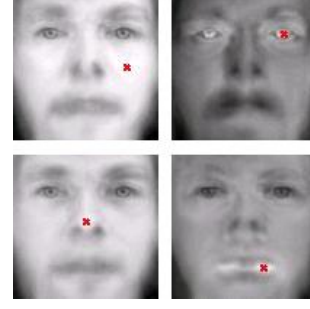


Figure 1: Correlation coefficient map (cheek, eye, nose, and mouth).

In the figure, we can see that the nearby pixels from the reference pixel are highly related with the reference pixels. This shows that the local information can be utilized in estimating value of the reference pixel.

On the other hand, looking at the correlation coefficient map of an eye, we can see that the pixels in the left eye are highly related to the one in the right eye. This shows that in addition to the local pixels, far away pixels from the reference point can also provide important information in estimating the pixel value of the reference point.

3.2 Jointly Gaussian distribution

In this part, we present a way to estimate an unknown pixel value from the other known ones.

Assume that the pixel value of x_i is not known, while that of x_j is known. We also assumed that pixel i and j follows jointly Gaussian distribution. Then, the pixel value of x_i can be obtained using a set of pixels $S_i = \{j \mid \text{index of } m \text{ number of } x_j \text{ which are highly related to } x_i\}$ that are highly related to x_i .

When $|S_i| = 1$ and the j^{th} pixel value is v , the conditional probability of x_i given x_j is

$$P(x_i | x_j = v) = \frac{P(x_i, x_j = v)}{P(x_j = v)}. \quad (11)$$

Because $P(x_j = v)$ is a constant, $P(x_i | x_j = v)$ is proportional to $P(x_i, x_j = v)$. With the assumption that the pixel i and j follows jointly Gaussian distribution, $P(x_i, x_j = v)$ becomes

$$\begin{aligned} P(x_i, x_j = v) &\propto \exp\left(-\frac{1}{2} \begin{pmatrix} x_i \\ v \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} x_i \\ v \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2} \frac{1}{(1 - \rho_{ij}^2) \sigma_i^2} \left(x_i - \frac{\sigma_i}{\sigma_j} \rho_{ij} v\right)^2\right) \end{aligned} \quad (12)$$

Here, Σ^{-1} is an inverse of the covariance matrix Σ of x_i and x_j .

$$\Sigma^{-1} = \frac{1}{(1 - \rho_{ij}^2)\sigma_i^2\sigma_j^2} \begin{bmatrix} \sigma_i^2 & -\rho_{ij}\sigma_i\sigma_j \\ -\rho_{ij}\sigma_i\sigma_j & \sigma_j^2 \end{bmatrix} \quad (13)$$

Using (12), the conditional mean and the variance of x_i are as follows:

$$E[x_i|x_j = v] = \frac{\sigma_i}{\sigma_j} \rho_{ij} v \quad (14)$$

$$\text{Var}[x_i|x_j = v] = (1 - \rho_{ij}^2)\sigma_i^2. \quad (15)$$

In case of $|S_i| = 2$ and assuming that the value of x_j is v and x_k is ω , we can estimate the value of x_i using a weighted sum of $E[x_i|x_j = v]$ and $E[x_i|x_k = \omega]$ because $E[x_i|x_j = v, x_k = \omega]$ is hard to represent.

$$E[x_i|x_j = v, x_k = \omega] \approx \alpha E[x_i|x_j = v] + \beta E[x_i|x_k = \omega] \quad (16)$$

In the above equation α and β should meet the condition $\alpha + \beta = 1$. The weight α can be regarded as the contribution of x_j in estimating the value of x_i . If the conditional variance of x_i given x_j decreases, the confidence of the estimation grows and the weight α should be increased. Therefore, the weight is set to be proportional to 1 divided by the standard deviation which is the square root of (15) in this paper. For example, when $|S_i| = 2$, $\alpha \propto 1/\sqrt{1 - \rho_{ij}^2}$, $\beta \propto 1/\sqrt{1 - \rho_{ik}^2}$ and it is made that $\alpha + \beta = 1$.

In case of $|S_i| > 2$, the weights are set in the similar way. If x_i and x_j are highly correlated to each other, we can assume high confidence resulting in a high weight. On the contrary, if the correlation between x_i and x_j are small, the corresponding weight is set to be small. In this paper, to alleviate computational complexity, we only make use of the pixels that have the highest correlation coefficient in estimating the unknown pixel value.

3.3 Detection and Recovery of Occluded Face Images

This paper shows how occluded face image is detected and recovered using correlation between pixels.

Obtaining correlation coefficient: The first step is to get the correlation coefficient for all the pixels

using the non-occluded training face images X_{tr} . This is done by using (10).

After getting the correlation coefficient of all the pairs of pixels, for each pixel, a list of highest correlation coefficient are stored.

Reconstruction: Assume that we do not know the first pixel value of the test face image which are centered by the mean image of training data. The pixel value can be predicted using a set of pixels which are highly correlated with the one in question.

The method to calculate each pixel value x_i with the condition $|S_i| = 2$ is shown from (14) to (16). This is generalized to a case of $|S_i| = m$ as follows:

$$w_j = \frac{1}{\sqrt{1 - \rho_{ij}^2}} \quad (17)$$

$$z_i = w_1 \frac{\sigma_i}{\sigma_1} \rho_{i1} x_1 + \dots + w_m \frac{\sigma_i}{\sigma_m} \rho_{im} x_m = \sum_{j=1}^m w_j \frac{\sigma_i}{\sigma_j} \rho_{ij} x_j \quad (18)$$

$$x'_i = \bar{x} + \frac{z_i}{\sum_{j=1}^m w_j} \quad (19)$$

Here, w_j means weight (or confidence) of x_j derived from the correlation coefficient between i^{th} and j^{th} pixels. The pixels x_j ($j = 1, \dots, m$) are the pixels that have the highest correlation coefficient with the i^{th} pixel.

Figure 2 illustrates the main idea of the proposed method. In the figure, \square mark corresponds to the pixel which needs to be estimated and \circ marks are the pixels highly correlated to the \square pixel. It is illustrated that the pixel value of \square is determined by the sums of weighted pixel values of \circ 's. By applying equations (17) to (19) to each of the pixel, we can obtain a new image x' .

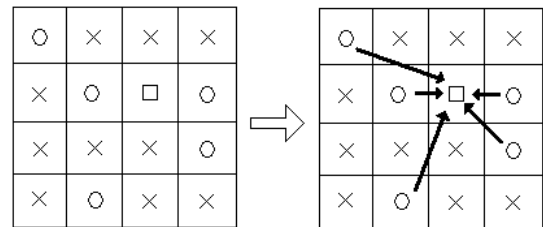


Figure 2: The image simply represented the part of a face image (a) and the image x_i is influenced by the pixels which have with correlation with x_i (b).

Occlusion detection: Once we obtain x' using the correlation coefficients, we can detect whether a pixel belongs to the occluded parts or not. The procedure is identical to the one used in the conventional PCA based method described in Section 2. Here, to clearly show how to detect occluded pixels we rewrite (7) in the following

$$\begin{aligned} |x'_i - x_i| < \epsilon &\Rightarrow i \in O^c \\ |x'_i - x_i| \geq \epsilon &\Rightarrow i \in O. \end{aligned} \quad (20)$$

Recovery: After detecting the occluded area from the entire image using (20), the occluded area should be filled with the estimated value. This is done by replacing the occluded pixel by the one in x' . The procedure is identical to the recovery step in Section 2 and (8) is rewritten here.

$$x''_i = \begin{cases} x_i & \text{if } i \in O^c \\ x'_i & \text{if } i \in O \end{cases} \quad (21)$$

After the recovery stage, x is replaced with x'' and the process is repeated until the difference between x and x'' becomes small enough. The final recovered image is obtained at the end of iteration when the process converges.

4 EXPERIMENTS

4.1 BioID Data

BioID data contains 1521 grayscale face images. It consists of frontal face images of 23 people and the size of each image is 100 by 100. The face images are fixed by the centers of eyes and the light condition is variable because the data were obtained in a real space.

Among 1521 images 1000 images were used for training and the other 521 images were used for testing.

We covered each test image with a squared box which has a random height and width on a random position to make an occlusion. Some examples of the occluded test images are shown on Figure 3.

4.1.1 Conventional PCA Based Method

We reduce a 10000 dimensional input space to a 50 dimensional space for the 1000 training images and make a weight matrix $W \in \mathbb{R}^{10000 \times 50}$.

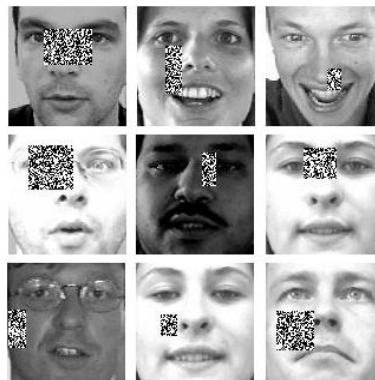


Figure 3: Occluded face images for BioID data.

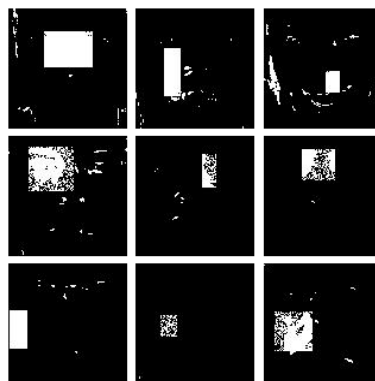


Figure 4: Occlusion detection for BioID data using the conventional PCA based method.



Figure 5: Recovered face images for BioID data using the conventional PCA based method.

Then we iteratively applied the PCA-based recovery algorithm in Section 2 and obtained the recovered image. Figure 4 shows the difference of pixel values between the original occluded image and the recovered image. We represent the occluded parts as value 1 and the non-occluded parts as value 0 on Figure 4. Final recovered images can be seen in

Figure 5. As a result of the detection and recovery, we can see a lot of noise in the figures.

4.1.2 The Proposed Method

We calculated correlation coefficients for 1000 training images using (10) and for each pixel, stored a number of pixel indices that have large correlation coefficient.

In a test stage, we can detect occluded parts by assuming jointly Gaussian distribution among pixel values. For simple implementation, for each pixel, we make use of other pixels that have correlation coefficients larger than 0.95.

Figure 6, shows the detected occluded area. In the figure, we can see some noise around the occluded parts, but the noise far from and within the occluded parts is not high compared to Figure 4.

Figure 7 shows the recovered image. As a result of recovery, we can see the continuity between the occluded parts and the non-occluded parts is enhanced resulting in smoother images than those in Figure 5.

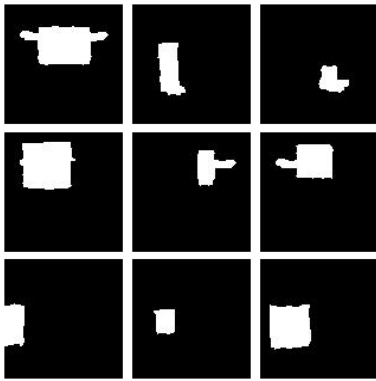


Figure 6: Occlusion detection for BioID data using the proposed method.



Figure 7: Recovered face images for BioID data using the proposed method.

4.1.3 Numerical Comparison

In this part, the performances of the recovery algorithms are numerically compared. In doing so, we define the mean absolute error and the mean squared error as follows:

$$\text{error}_{L1} = \frac{\sum_{i=1}^L \sum_{j=1}^P |x_{ji}^{\text{gt}} - x_{ji}^{\text{r}}|}{L \cdot P} \quad (22)$$

$$\text{error}_{L2} = \frac{\sum_{i=1}^L \sqrt{\sum_{j=1}^P (x_{ji}^{\text{gt}} - x_{ji}^{\text{r}})^2}}{L \cdot P} \quad (23)$$

Here, to L is the number of test image and P is the number of pixels in an image. x_{ji}^{gt} and x_{ji}^{r} represent the j^{th} pixel of the i^{th} ground truth test image and recovered test image respectively. The ground truth image is the one with no occlusion.

The errors are shown on the Table 1. As we expected, the proposed method showed less error than the conventional PCA-based method.

In Table 2, the number of iterations and the processing time of each method are shown. Both the number of iterations and the processing time of the proposed method are more than those of the PCA-based method. The reason can be attributed to the fact that the conventional PCA based method reconstructs the image at once by multiplying the weight matrix W , while in the proposed method reconstruction is done pixel by pixel.

4.2 Yale Data

The Yale data contains 165 grayscale images of 100 by 80 in size. It consists of 11 images per 15 individuals in variable light conditions.

Among the 165 images, 100 images were used for training and the remaining 65 were used for testing.

Table 1: MAE and MSE between recovered BioID data and non-occluded BioID data.

	Absolute error	Mean-squared error
PCA based	17052.45	788.06
Proposed	10500.65	532.22

Table 2: Number of iterations and processing time for BioID data.

	# of iteration		Processing time	
	Avg.	Var.	Avg.	Var.
PCA based	3.117	1.226	0.092	0.051
Proposed	5.633	3.820	1.173	0.421

As the experiments on BioID data, the test images were intentionally occluded as shown in Figure 8.

4.2.1 Conventional PCA Based Method

The 8000 dimensional input space is reduced to a 50 dimensional subspace by multiplying the weight matrix $W \in \mathbb{R}^{8000 \times 50}$ obtained from the 100 training images.



Figure 8: Occluded face images for Yale data.

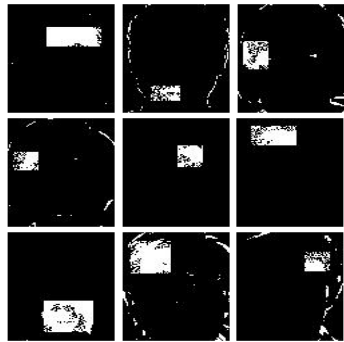


Figure 9: Occlusion detection for Yale data using the conventional PCA based method.



Figure 10: Recovery of occluded face images for Yale data using the conventional PCA based method.

Figure 9 and Figure 10 show the detected occlusion area and the recovered image respectively. As in the BioID experiment, the detected occlusion area contains a lot of noise in both the occluded and the non-occluded parts. The recovered images also show a blurring phenomenon and discontinuity as in the BioID case.

4.2.2 The Proposed Method

Figure 11 and Figure 12 are the detected occlusion areas and the recovered images by the proposed method respectively. We can see that the detected occlusion area is more continuous with much less noise than that of the PCA-based method. The recovered images have less blurs and much more smoothly connected in the boundary between the occluded parts and the non-occluded parts.

4.2.3 Numerical Comparison

Table 3 is the computed mean absolute error and the mean squared error using (22) and (23). As in Table 1, the error values of the proposed method are smaller than those of the conventional PCA based method.

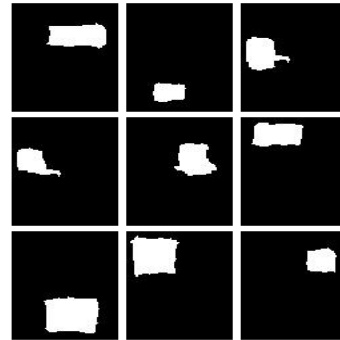


Figure 11: Occlusion detection for Yale data using the proposed method.



Figure 12: Recovery of occluded face images for Yale data using the proposed method.

Table 3: MAE and MSE between recovered Yale data and non-occluded Yale data.

	Absolute error	Mean-squared error
PCA based	32389.53	1416.22
Proposed	16142.30	840.42

Table 4: Number of iterations and processing time for Yale data.

	# of iteration		Processing time	
	Avg.	Var.	Avg.	Var.
PCA based	4.908	1.234	0.017	0.021
Proposed	5.6	3.432	1.403	0.749

That means the recovered images by the proposed method are much similar to the original Yale images than those by the PCA-based method.

Table 4 shows the number of iterations and processing time of each method. As in BioID case, both the number of iterations and the processing time were better in the conventional PCA-based method than the proposed method.

5 CONCLUSIONS

In this paper, we proposed a new method to recover the occluded face images using the correlation coefficient between pairs of pixels. The simple idea that a pixel value can be determined by the weighted sum of other pixel values which are highly correlated with the one in question.

The proposed method is compared with the conventional PCA based method and it showed better recovery performance in both qualitatively and quantitatively. The blurring of the recovered images is much less and the border lines between occluded and non-occluded parts are connected well. Moreover, the mean absolute error and the mean squared error value of the proposed method are smaller than the PCA-based method. However, the proposed method was comparatively slower than the PCA-based method and this should be enhanced in the future work.

ACKNOWLEDGEMENTS

This work was supported by Korea Research Foundation Grant funded by Korean Government (KRF-2011-0005324).

REFERENCES

- Wu, C., Liu, C., Shum, H., Y., Xy, Y., Q., Zhang, Z., 2004. "Automatic eyeglasses removal from face images". *Asian Conference on Computer Vision, IEEE Trans. Pattern Analysis and Machine Intelligence.*, Vol. 26, no. 3, pp. 322-336.
- Wang, Z., M., Tao, J., H., 2007. "Reconstruction of partially occluded face by fast recursive PCA". *International Conference on Computational Intelligence and Security Workshops.*, pp. 304-307.
- XiaoFeng, S., Min, T., deren, C., Ruofeng, T., 2010. "A two step method to recover occluded part of face". *International Conference on Information Sciences and Interaction Sciences.*, pp. 169-173.
- Hwang, B., W., Lee, S., W., 2003. "Reconstruction of partially damaged face images based on a morphable face model". *IEEE Trans. Pattern Analysis and Machine Intelligence.*, Vol. 25, no. 3, pp. 365-372.
- Lin, D., Tang, X., 2007. "Quality-driven face occlusion detection and recovery". *IEEE Conference of Computer Vision and Pattern Recognition.*, pp. 1-7.
- Efros, A., A., Leung, T., K., 1999. "Texture synthesis by non-parametric sampling". *IEEE International Conference on Computer Vision.*
- Kwatra, V., 2003. "Graphcut textures: Image and video synthesis using graph cuts". *SIGGRAPH.*
- Jia, J., Tang, C.,-K., 2003. "Image repairing: Robust image synthesis by adaptive and tensor voting". *CVPR.*
- Sun, J., Yuan, L., Jia, J., Shum, H.,-Y., 2005. "Image Completion with Structure Propagation". *SIGGRAPH.*
- Komodakis, N., Tziritas, G., 2006. "Image Completion Using Global Optimization". *IEEE CVPR.*