Occlusion Images Generation from Occlusion-Free Images for Criminals Identification based on Artificial Intelligence Using Image

Yonggeol LEE*, JangMook KANG**

*Police Science Institute, Asan, Korea
** Department of Big-Data Industrial Security, Namseoul University, Cheonan city 31020, Korea,
**Corresponding author E-mail:honukang@gmail.com; kangjm@nsu.ac.kr

Abstract

In this paper, we propose the method to generate occlusion images which include sunglass and scarf from occlusion-free images for criminals' identification. First, we construct an auxiliary set that has occlusion-free face image and its occluded face images from AR database. Secondly, pixel weights are determined by the statistical analysis of the degree depending on whether occlusion-free or occlusion. As the weight is closer to 1, the pixel value of original image is maintained as much as possible. Therefore, pixel values of occlusion-free region are preserved when the weights are closed to 1. Third, pixel values of the occlusion region of original image are replaced to pixel values which are the mean of occlusion images by multiplying the composited weights. As a result of generate face images with various occlusions using the proposed method, the images were able to preserve the person-specific characteristic, and criminal's identification has become easier.

Keywords: *Image Synthesis, Criminals Identification, Pixel Weights, Statistic*

1. Introduction

Recently, face recognition systems are increasingly used in various fields[1]. In particular, the police actively use video devices such as intelligent CCTV, smart phone and dash camera to arrange for a quick arrest. However, occlusion, illumination, and camera angle variation act as a factor to reduce performance when facial recognition is acquired from images acquired from these devices.

On the police side, it is difficult to arrest the criminals who disguise to hide their identity by wearing accessories, generally. Police officers store occlusion-free images in the database and only consider various shooting angles about the arrested criminals. In the future, if a criminal committed a crime, the suspect can be arrested promptly by comparing and analyzing the face photograph of the suspect taken on the CCTV and the face photograph of the criminal constructed in the database. However, immediately after the crime, most criminals wear accessories such as a sunglass, scarf, wig and hat to hide their identity[2]. Therefore, even if criminals are included on the CCTV, it is difficult to judge whether they match the criminals of the existing database easily.

A fundamental way of solving this problem, it is to establish a database of many criminals that contain various variations. However, it is impossible to respond to the various variations present in real-world[3]. Therefore, a method for synthesizing virtual images from occlusion-free images is needed.

In this paper, we propose the method to generate occlusion images which include sunglass and scarf from occlusion-free images for criminals' identification. First, we construct an auxiliary set that has occlusion-free face image and its occluded face images from AR database[4]. Secondly, pixel weights are determined by the statistical analysis of the degree depending on whether occlusion-free or occlusion[5]. As the weight is closer to 1, the pixel value of original image is maintained as much as possible. Therefore, pixel values of occlusion-free region are preserved when the weights are closed to 1. Third, pixel values of the occlusion region of original image are replaced to pixel values which are the mean of occlusion images by multiplying the composited weights.

As a result of generate face images with various occlusions using the proposed method, the images were able to preserve the person-specific characteristic, and criminal's identification has become easier.

2. Proposed Methods

In this paper, we propose an image synthesis method using statistical features which are different for each pixel position due to occluders' change.
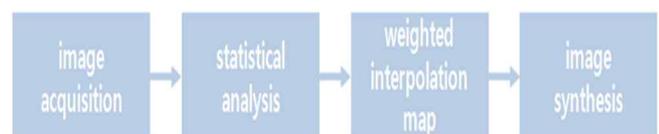

Fig. 1: The flow chart for image synthesis

First, collect the occlusion-free image and the corresponding occlusion image with the sunglass and scarf. Secondly, we analyse the degree of change according to occludes statistically from collected images, and define a pixel weights which is a set of weights

to be applied to each pixel position from this statistical analysis. Finally, occlusion image generation using input image and pixel weights. (Fig. 1)

2.1 Auxiliary Set

In this paper, we compose auxiliary set as follows. Auxiliary set contains 150 images of 50 classes(25 men and 25 women) from AR database which include occlusions such as occlusion-free, sunglass and scarf. All the images were aligned to a size of r (pixels) $\times c$ (pixels) and converted d-dimensional column vectors.

Table 1: Summary of variables, vectors, and matrices

symbol	dimension	definition
r	1	image width
c	1	image height
d	1	dimension of image($= r \times c$)
x_i	$\Re^d$	occlusion-free image($i = 1, \cdots, n$)
y_i	$\Re^d$	occlusion image corresponded to x_i ($i = 1, \cdots, n$)
X	$\Re^{d \times n}$	auxiliary set of occlusion-free images
Y	$\Re^{d \times n}$	auxiliary set of occlusion images
$\overline{Y}$	$\Re^{d \times 1}$	mean image of Y
λ	$\Re^{d \times 1}$	standard deviation
λ^{norm}	$\Re^{d \times d}$	normalized weights($\in [0,1]$)
I_q	$\Re^d$	query(input) image
z	$\Re^d$	synthesis image

In Table 1, set of occlusion-free images $X = [x_1, x_2, \cdots x_n]$ and set of occlusion images $Y = [y_1, y_2, \cdots y_n]$ are composed occlusion-free image x_i and occlusion image y_i corresponded to x_i.

2.2. Pixel Weighs

If there are occlusion-free and occlusion images that is taken under strictly constrained facial expressions or poses, the degree of change in pixels at the same position between two images is determined by occlusions. Fig. 2 shows the difference in pixel values according to occlusion at a particular position when changing from x_i to y_i. The difference surrounding eyes is remarkable when wearing a sunglass. Similarly, the difference surrounding mouth is noticeable when wearing a scarf.

Occlusion images are reconstructed by using difference ($\in [-255, 255]$) between occlusion-free and occlusion images. On the other hand, occlusion-free images are reconstructed by using difference between occlusion and occlusion-free images, also.

In real-world, difference between occlusion-free and occlusion images cannot be acquired for all person. Therefore, it needs statistical analysis to find general patterns between occlusion-free and occlusion images. To solve this problem, the object function is defined as follows[5].

$$\arg\min_{\lambda \neq 0} \| y_i - \{\lambda x_i + (1 - \lambda) y_i\} \|, \quad (i = 1, \cdots, n), \quad (1)$$
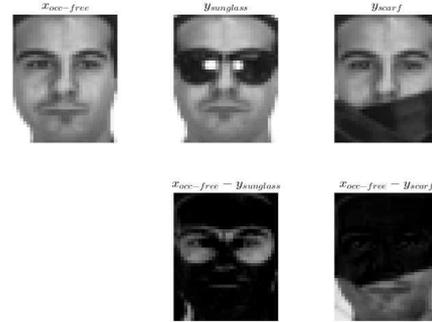

Fig. 2: Used images in AR database. (1st row) occlusion-free image and occlusion image with sunglass and scarf (2nd row) difference between occlusion-free and occlusion images

The goal is to find pixel weights λ ($\Re^{d \times 1}$) that can minimize errors in the objective function for all data. λ represents the degree of change in pixels at the same position between occlusion-free and occlusion images[4]. λ is defined as follows using the log-scaled standard deviation of the absolute difference.

$$\lambda = \log_2(std(| X - Y |)), \quad (2)$$

In addition, Equation 2 can be redefined as follows.

$$\lambda = \log_2 \left(\frac{1}{n-1} \sqrt{\sum_{i=1}^{n} (| x_i - y_i | - \mu)} \right), \quad (3,1)$$

$$\mu = \frac{1}{n} \sum_{i=1}^{n} | x_i - y_i |, \quad (3.2)$$

In Equation 3, the absolute difference means difference of $X \cap Y$ and $X \cup Y$, The standard deviation expresses the degree of statistical variation of the entire image from the difference between pixels of these individual images. A high value is obtained for pixels with large variation and a low value is obtained for pixels with small variation such as accessories.
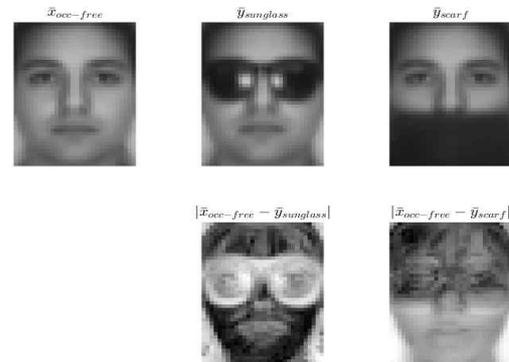
Fig. 3: The mean images. (1st row) occlusion-free image and occlusion image with sunglass and scarf (2nd row) standard deviation of absolute difference between occlusion-free and occlusion images

In Fig. 3, occlusion-free image and occlusion images, respectively. In the case of sunglass, a strong variation occurred around the eyes, and the degree of change in the mouth and nose was severe in the scarf. On the contrary, it can be confirmed that there is almost no variation in the nose area and the intercuspal area.

In Equation 3, when the value of the pixel is very large according to the degree of the variation of the pixel, there is a saturated pixel whose pixel value exceeds 255 in the synthesized image. In addition, it is necessary to have a range of $[0,1]$ in the object function, since it determines to which weight among the two given images to apply more weights. Therefore, the normalized λ^{norm} is calculated as follows.

$$\lambda^{norm} = \frac{\lambda - \min(\lambda)}{\max(\lambda) - \min(\lambda)}, \tag{4}$$

λ^{norm} is d-dimensional vector from Equation 4. For this, we can apply a weight to each pixel of an image composed of dimensional column vectors by converting it into a diagonal matrix of size $d \times d$.

$$\lambda^{norm} = \begin{cases} 0 \leq \lambda^{norm}_{ij} \leq 1, & i = j \\ 0 & otherwise \end{cases}, \tag{5}$$

2.3. Synthesis Images

To generate the synthesis image z from query(input) image I_q as follows.

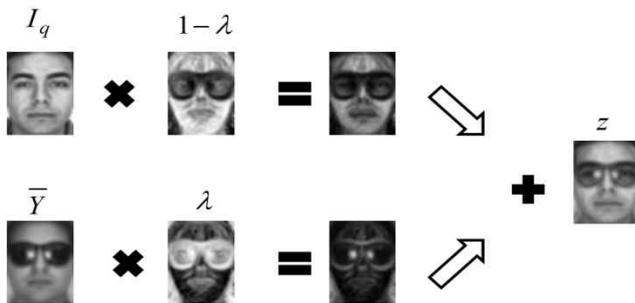$$z = (1 - \lambda^{norm})I_q + (\lambda^{norm})\overline{Y}, , \tag{6}$$

Fig. 4: The schematic diagram for image synthesis.

In Equation 6, the collected average image $\overline{Y}$ and λ^{norm} generated using z reflect the fine facial deformation and have the advantage of preserving individual information such as glasses, spots, and pimple (Fig. 4).

3. Experimental Results

AR database contains over 4,000 color images corresponding to 126 people's faces (70 men and 56 women)[4]. Images feature frontal view faces with different facial expressions, illumination conditions, and occlusions such as sunglass and scarf. Each person participated in two sessions, separated by two weeks (14 days) time.

	auxiliary set	recognition set	
	session 1 + session 2	training set (session 1)	test set (session 2)
No. of class	50	50	50
No. of images per class	3	1	3
variations	occlusion-free, sunglass, scarf	occlusion-free	occlusion-free, sunglass, scarf

The same pictures were taken in both sessions. In this paper, we use 600 gray images of 100 classes(50 men and 50 women) which had already aligned to a size of 42 (pixels) $\times$ 30 (pixels)[6]. We divide the auxiliary set(25 men and 25 women) to generate λ and the recognition set to face recognition experiment.
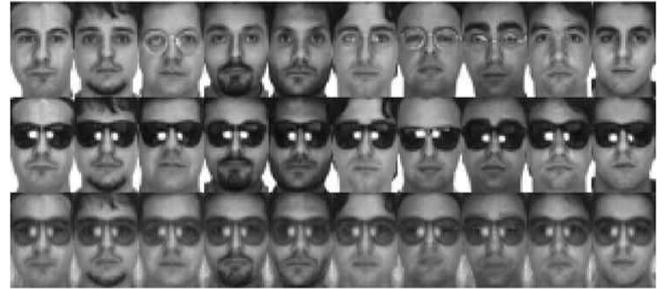

Fig. 5: Compared with occlusion-free images, occlusion images corresponded occlusion-free images and synthesis images. (1st row) occlusion-free images (2nd row) occlusion images (3rd row) synthesis images with sunglass
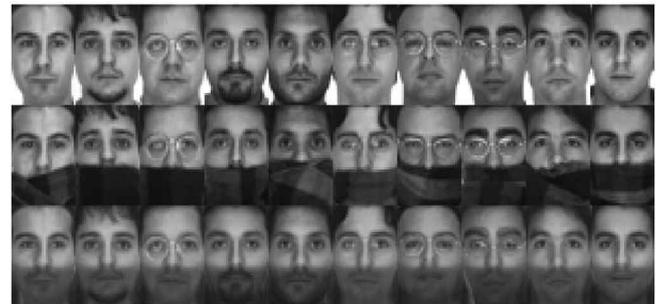

Fig. 6: Compared with occlusion-free images, occlusion images corresponded occlusion-free images and synthesis images. (1st row) occlusion-free images (2nd row) occlusion images (3rd row) synthesis images with scarf

Fig. 5 shows an occlusion-free image in the AR database, an occlusion image wearing sunglasses, and a composite image generated using λ. The synthesis image represents the average shape and position of the images wearing sunglasses, and even reflects up to glare. It also preserves person-specific information such as the beard and hair style. Fig. 6 shows the synthesis image generated for the scarf. Unlike sunglasses, scarves are not represented in the form of average scarf images, since they vary in color, wearing style, and so on. However, occlusion occurred around the mouth in the generated image, and it can be confirmed that person-specific information such as beard and hair style is preserved as in fig. 5.
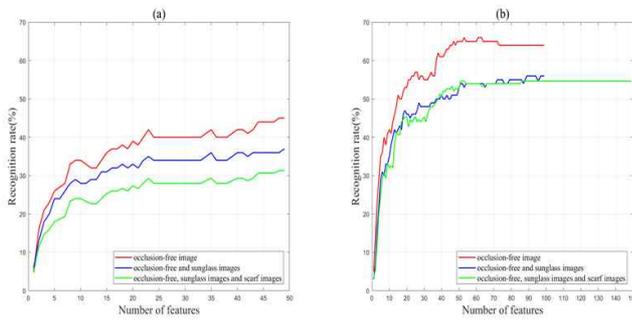
Fig. 7. Face recognition result according to the number of features. (a) using only occlusion-free (b) using occlusion-free and synthesis images from proposed method

In this paper, we use Principal Component Analysis(PCA)[7] among traditional appearance-based face recognition methods[8-11], to verify the performance of the proposed method. Fig. 7 shows the face recognition result of the test set according to the image used for PCA learning. Fig. 7 (a), we used a single occlusion-free image for each individual in the training set, resulting in less than 50% face recognition for all cases. Fig. 7 (b), the occlusion-free images used in (a) and the corresponding images for occlusion-free images were used together. Experimental results show that face recognition performance is significantly improved for all variations when using synthesis images.

training set (each 50 images)	result for subset of test set (each 50 images)			total result
	occlusion-free	sunglass	scarf	
occlusion-free images	70.0%	20.0%	4.0%	31.3%
Proposed method (occlusion-free and synthesis images)	80.0%	54.0%	30.0%	54.7%

Experimental results show that the proposed method improves the face recognition performance by about 23.4% over the experiment using only the occlusion-free image. More specifically, the differences were 10.0%, 34.0% and 26.0% for occlusion-free, sunglass and scarf, respectively. For scarf, only 2 of the 50 were recognized. since they vary in color, wearing style, and so on (Fig. 6)

4. Conclusion

There are two reasons why it is difficult for a police officer to arrest a criminal by utilizing a face recognition system. First, criminals use various accessories such as sunglasses, scarves, and hats to hide themselves as much as possible. In general, these occluders lead to poor face recognition performance. More serious is the frontal occlusion-free images, which are stored only when the officer registers criminal image information, since only the shooting angle is considered. Therefore, given the disguised criminal image collected from the CCTV or dash camera, the face recognition result indicated by the face recognition system becomes unreliable.

In this paper, we propose a face composition method to improve facial recognition performance in a limited situation where only occlusion-free images exist. From a statistical point of view, there is a significant difference between certain occlusion-free images and occlusion-free images compared to occlusion-free images. For example, wearing sunglasses increases the difference in pixels around the eyes. Therefore, we construct an auxiliary set that has occlusion-free face image and its occluded face images from the AR database and extract weights to apply to individual pixels from the standard deviation of the absolute difference. To generate occlusion image, pixel weights used to an occlusion-free image and the mean of the occlusion image of auxiliary set. The synthesized

images preserve the person-specific characteristics such as beard and hair style.

Experimental results show that only occlusion-free images are vulnerable to occlusion images when used in face recognition models. When a synthetic image is generated using the proposed method and used for face recognition model learning, the performance is improved by about 23% compared to the existing performance.

Acknowledgement

References

[1] Choi, S. I., Lee, Y., & Kim, C. (2015). Confidence measure using composite features for eye detection in a face recognition system. IEEE Signal Processing Letters, 22(2), 2015, 225-228.

[2] Demleitner, Nora V. Witness Protection in Criminal Cases: Anonymity, Disguise or Other Options?. The American Journal of Comparative Law 46.suppl_1 (1998): 641-664.

[3] Kim, Chunghoon, et al. A new biased discriminant analysis using composite vectors for eye detection. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 42.4 (2012): 1095-1106.

[4] Martinez, A. M. (1998). The AR face database. CVC Technical Report24, 1998.

[5] Choi, S. I., Eom, T. K. Lee, Y., Facial Expression Synthesis Using Normalized Pixel Weights Between Expressions, Conference on The Institute of Electronics and Information Engineers, 6(2015), 664-666.

[6] Gao, Guangwei, et al. Robust face recognition via multi-scale patch-based matrix regression. PloS one 11.8 (2016): e0159945.

[7] Turk, M. A., & Pentland, A. P. Face recognition using eigenfaces. In Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on , 1991, pp. 586-591.

[8] Belhumeur, Peter N., João P. Hespanha, and David J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. Yale University New Haven United States, 1997.

[9] Cevikalp, Hakan, et al. Discriminative common vectors for face recognition. IEEE Transactions on pattern analysis and machine intelligence, 27.1 (2005): 4-13.

[10] Yu, Hua, and Jie Yang. A direct LDA algorithm for high-dimensional data-with application to face recognition. Pattern recognition 34.10 (2001): 2067-2070.

[11] Kim, Chunghoon, and Chong-Ho Choi. "Image covariance-based subspace method for face recognition." Pattern recognition 40.5 (2007): 1592-1604.