Facial Occlusion Detection via Structural Error Metrics and Clustering

Xiao-Xin Li[1(✉)], Ronghua Liang[2], Jiaquan Gao[1], and Haixia Wang[2]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China
{mordekai,gaojq}@zjut.edu.cn

[2] College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
{rhliang,hxwang}@zjut.edu.cn

Abstract. Facial occlusions pose significant obstacles for robust face recognition in real-world applications. To eliminate the effect incurred by occlusions, most of the popular methods concentrate on dealing with the error between the occluded image and its recovery. Inspired by the working mechanism of human visual systems in facial occlusion detection, we suggest that it should be the error metric and clustering rather than exact recovery that play important roles for occlusion detection. By considering the structural differences between faces and occlusions, such as colors and textures, we construct five structural error metrics. By considering the common structures shared by all occlusions, such as localization and contiguity, we construct a structured clustering operator. Furthermore, we select the optimal error metric via the minimum occlusion boundary regularity criterion. Integrating the above techniques, we propose the Structural Error Metrics and Clustering (SEMC) algorithm for facial occlusion detection. Experimental results demonstrate that, even just using the mean face of the training images as the recovery image, SEMC still achieves more accurate and robust performance compared to the related state-of-the-art methods.

Keywords: Unconstrained face recognition · Facial occlusion detection · Structural error metric · Structural clustering

1 Introduction

Recently, recognizing human faces with occlusions has received a lot of attention in computer vision and pattern recognition [4,7,11,13,15,16]. Facial occlusions, including accessories, shadows or other objects in front of a face, pose significant obstacles for robust face recognition in real-world applications [3]. To eliminate the effect incurred by occlusion, researchers have studied the solution schemes from different views. Most of these schemes concentrate on the recovery error $\hat{e} \in \mathbb{R}^m$ between the occluded image $y \in \mathbb{R}^m$ and its recovery $\hat{y} \in \mathbb{R}^m$ with respect to (w.r.t.) the training dictionary $A \in \mathbb{R}^{m \times n}$, with the assumption that the

© Springer International Publishing Switzerland 2015
X. He et al. (Eds.): IScIDE 2015, Part I, LNCS 9242, pp. 118–127, 2015.
DOI: 10.1007/978-3-319-23989-7_13

larger the entry values of the error $\hat{e}$, the higher the probability with which the corresponding pixels are occluded. By assuming that the error $\hat{e}$ can be sparsely coded w.r.t. some error coding dictionary $E \in \mathbb{R}^{m\times d}$, researchers proposed the following Error Coding Model (ECM) from different views [2,5,8,13,14]

$$\min_{x,c} \|x\|_1 + \|c\|_1 \quad s.t. \quad [A\ E] \begin{bmatrix} x \\ c \end{bmatrix} = \hat{y} + \hat{e} = y, \tag{1}$$

where $x \in \mathbb{R}^n$ and $c \in \mathbb{R}^d$ are the coding coefficients of the recovery image $\hat{y}$ and the error $\hat{e}$, respectively.

Another view on dealing with the error is the Error Weighting Model (EWM). According to the works of [4,7,15,16], the EWM can be summarized as

$$\min_{x,w} \|x\|_{\ell^a} + \mu \|w \odot \hat{e}\|_{\ell^b} + \lambda\phi(w, \hat{e}) \ s.t. \ \hat{e} = y - Ax, \tag{2}$$

where ℓ^a and ℓ^b are the norm indexes, μ and λ are the regularization parameters, $w \in \mathbb{R}^m$ is the error weight, $\phi(w, \hat{e})$ is the cost function of w w.r.t. the error $\hat{e}$, and $\odot$ is the Hadamard product. With $\mu = 1, \lambda = 0, w = 1, \ell^a = \ell^b = 1, E = I$, the EWM is equivalent to the ECM. The entries of the weight w indicate the occlusion support or the probabilities with which the corresponding pixels are occluded.

The occlusion location is mainly indicated by the reconstruction error $\hat{e} = y - \hat{y} = y - Ax$ in both ECM and EWM. It seems that the quality of the recovery image $\hat{y}$ determines the accuracy of the corresponding occlusion detection result. However, we know that an exact recovery does not be necessary for human visual system (HVS), that is, HVS could recognize the occluded region of a face without having to see this face before, i.e., without having to compare it with its ground truth. HVS captures the structural differences by comparing the occlusion with a *fuzzy* face model, which is learned from the faces emerging in everyday life ever before. We call that the learned face model is fuzzy as its identity is unclear and HVS only keeps its typical structure composed of eyes, nose, mouth and etc. Inspired by this observation, we explore the automatic facial occlusion detection technology via structural comparison, that is, structural error metrics and clustering. For simplicity, we choose the mean face $\bar{y} = \frac{1}{n}(A \times \mathbf{1})$ ($\mathbf{1} \in \mathbb{R}^n$) as the fuzzy face model in the subsequent work.

The rest of this paper is organized as follows. Section 2 presents several structural error metrics and a structural error clustering operator. Section 3 gives an optimal error metric selection criterion, which integrates the proposed structural error metrics and clustering operator together. Section 4 performs the experiments. Section 5 concludes the paper.

2 Structural Error Metrics and Clustering

In order to highlight the error incurred by occlusion, we measure the error $\tilde{e}$ between the test image y and the mean face $\bar{y}$ based on the potential differential structures between faces and occlusions; in order to cluster the occluded pixels, we consider the error $\tilde{e}$ and its clustering operator based on the common structures shared by all occlusions.

2.1 Structural Differences Induced Error Metrics

By observation, we note that the structural differences between occlusions and the fuzzy face model (the mean face) is mainly reflected in colors, textures and shapes. We now consider how to use these structural differences to construct error metrics. Suppose we could project an image into a structured subspace, which keeps or strengthens the preconceived structure and wipes off or weakens the unwanted structures. Then the error between two images can be measured in this new subspace

$$\tilde{e}_f = \mathcal{E}_f(y, \bar{y}) = |f(y) - f(\bar{y})|, \tag{3}$$

where the function $f : \mathbb{R}^m \to \mathbb{R}^m$ is the structural difference descriptor embedded in the structured subspace. We call the error metrics based on (3) the structural error metrics (SEMs). Hence, what is critical for SEMs is to design the structural difference descriptors.

Color Difference Metrics. The color difference can be directly measured by the absolute error metric in the original image domain $\mathcal{E}_I(y, \bar{y}) = |I(y) - I(\bar{y})| = |y - \bar{y}|$. However, $\mathcal{E}_I$ does not consider the relative error between image pixels. For example, if there exists $y_j = y_i + a$ and $\bar{y}_j = \bar{y}_i + a$ for some $a \gg 1$, we have $\mathcal{E}_I(y_i, \bar{y}_i) = \mathcal{E}_I(y_j, \bar{y}_j)$. While it seems reasonable in mathematics, this error measurement result clashes with *Weber's law* [12] in psychophysics, which says that the relationship between the stimulus S and the perception p is logarithmic. That is, the lower the initial stimulus is, the more easily it could be perceived. By assuming that the occlusion is a stimulus with low intensities, we project images into the logarithmic domain and have the log-based error metric $\mathcal{E}_{\log}(y, \bar{y}) = |\log y - \log \bar{y}|$, which enhances the error caused by the low-value occlusions while suppresses the error incurred by high-value occlusions.

We then consider occlusions with high intensities. This problem might be well solved, if we can map the occluded image into a feature subspace, where most of the pixels with high values are transformed to the ones with low values. By supposing the pixel values of the occlusion in the local area change slowly and smoothly, this feature subspace can be described by the gradient $\nabla I = \sqrt{\left(\frac{\partial I}{\partial x}\right)^2 + \left(\frac{\partial I}{\partial y}\right)^2}$, where $\frac{\partial I}{\partial x}$ and $\frac{\partial I}{\partial y}$ are the gradients along the vertical and horizontal directions, respectively. We now have the log-gradient-based error metric $\mathcal{E}_\nabla(y, \bar{y}) = |\log_\nabla(y) - \log_\nabla(\bar{y})|$, where $\log_\nabla(y) = \log(\nabla y)$.

Texture Difference Metrics. While the log-based error metric $\mathcal{E}_{\log}$ and the log-gradient-based error metric $\mathcal{E}_\nabla$ are sensitive to occlusions with low intensities and high intensities changing uniformly, they are insensitive to occlusions with intensities changing rapidly or randomly. In this scenario, it is the texture differences rather than the color differences that dominate the structural differences between faces and occlusions. However, texture is easy to see but difficult to define [10], as its definition might be different for different applications.

In this work, we describe the texture by the density of edges per unit area, which can be computed over an image area by the Laplacian filtering $\nabla^2 I = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2}$. We then have a new error metric in the Laplacian filtered domain $\mathcal{E}_{\nabla^2}(y, \bar{y}) = \left|\nabla^2 y - \nabla^2 \bar{y}\right|$, which we call the Laplace-based error metric.

Color-Texture-Combined Difference Metrics. We now consider the scenario when both the color and texture differences are prominent. Here, we introduce the differential excitation (DE) of an image proposed by Chen *et al.* [1] $DE(I) = \arctan \frac{\nabla^2 I}{I}$. The DE operator simultaneously keeps the texture and color features of the original image, as the Laplacian filtered image $\nabla^2 I$ in the numerator calculates the texture feature and I in the denominator keeps the color feature. Specifically, the ratio $\frac{\nabla^2 I}{I}$ actually amounts to the *Weber fraction*, and the arctangent function limits the output of DE in $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ and is a logarithm-like function, which is also sensitive to the change incurred by low intensities. We now have the DE-based error metric $\mathcal{E}_{DE}(y, \bar{y}) = |DE(y) - DE(\bar{y})|$.

2.2 Sharing Structures Induced Error Clustering

Clearly, it is critical to seek the common structures shared by all occlusions for clustering occluded pixels. The common structures of occlusions explored in existing literature mainly includes locality [4] and contiguity [16].

Local Structure for Error Enhancement and Normalization. The correntropy induced metric (CIM) [4] measures the error between each pixel pair y_i and $\bar{y}_i$ as follows $CIM(y_i, \bar{y}_i) = 1 - g(y_i - \bar{y}_i)$, where $g(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right)$ is the Gaussian kernel. CIM is a statistical local metric due to the utilized Gaussian kernel and its locality can be adjusted with the kernel size σ. In order to utilize the local structure of the error calculated by various error metrics, we extend the CIM to the following form $CIM_f(y_i, \bar{y}_i) = 1 - g(\mathcal{E}_f(y_i, \bar{y}_i))$. Note that the error metric LD proposed in [7] is just an instance of the error metric CIM_f.

Contiguous Structure for Error Clustering. The contiguous structure of occlusion is usually depicted by the adjacent relationships of the occlusion spatial support. However, the error support, instead of the occlusion support, is commonly used in literature, since the occlusion support is usually unknown. The contiguous filtering combined with the error clustering is the common way to obtain the error support. In [16], the authors explored the contiguous structure via Markov random field, and the error support is estimated by solving a binary GraphCut problem. In this work, for simplicity, we just use the morphological filtering [16] to obtain the contiguous structure. The key idea is to first cluster the errors by K-means (or just threshold the errors) and then to apply open and close operations to the binary error support, which can be formulated as $\mathcal{K}_s(\hat{e}) = f_\bullet(f_\circ(\mathcal{K}(\hat{e})))$, where $f_\circ(\cdot)$ and $f_\bullet(\cdot)$ are open and close operations, respectively.

Algorithm 1. Structural Error Metrics and Clustering (SEMC) for Facial Occlusion Detection

Input: data matrix $A \in \mathbb{R}^{m \times n}$, test sample $y \in \mathbb{R}^m$.
Output: detected occlusion support $\tilde{s}$.

1. Calculate the mean face $\bar{y} = \frac{1}{n}(A \times \mathbf{1})$;
2. Set the structure difference operator ensemble: $F = \{I, \log, \nabla, \nabla^2, DE\}$;
3. **For** each $f \in F$
4. Calculate the recovered error $\tilde{e}_f$: $(\tilde{e}_f)_i = CIM_f(y_i, \bar{y}_i)$;
5. Cluster the recovered error: $\tilde{s}_f = \mathcal{K}_s(\tilde{e}_f)$;
6. **End For**
7. Select the optimal occlusion support: $\tilde{s} = \tilde{s}_{f^*}$, where $f^* = \arg\min_{f \in F} \mathcal{B}(\tilde{s}_f)$.

3 Optimal Error Metric Selection

We now have 5 structural error metrics, $\mathcal{E}_I$, $\mathcal{E}_{\log}$, $\mathcal{E}_{\log_\nabla}$, $\mathcal{E}_{\nabla^2}$ and $\mathcal{E}_{DE}$, for facial occlusion detection, which are designed for different occlusions with different structures, respectively. As the structure of a special occlusion is usually priori unknown, it seems difficult to automatically choose the optimal error metric. However, we find that the minimum occlusion boundary regularity criterion proposed in [7] can be used here to help selecting the optimal error metric. The idea is inspired by the observation that all natural occlusions usually have smooth and regular boundaries. We therefore deduce that if the shape of the detected occlusion based on an error metric is coarse and irregular, the corresponding utilized error metric might not be the optimal one. According to the morphological boundary detection algorithm presented in [7], the minimum boundary regularity criterion can be formulated as $\arg\min_f \mathcal{B}(s_f) = \|s_f - (s_f \boxminus T)\|_1$, where s_f is the detected error support, $\boxminus$ is the erosion operator, and $T = [1\,1\,1; 1\,1\,1; 1\,1\,1]$ is the structuring element.

Incorporating the 5 structural error metrics, $\mathcal{E}_I$, $\mathcal{E}_{\log}$, $\mathcal{E}_{\log_\nabla}$, $\mathcal{E}_{\nabla^2}$ and $\mathcal{E}_{DE}$ with the local error metric CIM, and using the structured clustering operator $\mathcal{K}_s$ and the minimum occlusion boundary regularity criterion, Algorithm 1, dubbed the Structural Error Metrics and Clustering (SEMC), summarizes the whole procedure of our method used to make facial occlusion detection.

4 Simulations and Experiments

To evaluate the proposed SEMC algorithm, we compare it with the state-of-the-art methods on two publicly available databases, namely, the Extended Yale B [6] database and the AR [9] database. Since the ECM (1) does not contain the occlusion detection mechanism, we just pay attention to the state-of-the-art EWM-based methods: the correntropy-based sparse representation (CESR) [4], the robust sparse coding (RSC) [15], and the structured sparse error coding (SSEC) [7]. Note that both CESR and RSC just calculate the probability w with

which the pixels are occluded. We therefore cluster w to estimate the occlusion support: for CESR, we estimate the occlusion support by K-means clustering $s_{CESR} = \mathcal{K}(1-w)$; for RSC, according to its open Matlab code, we estimate the occlusion support by threshold clustering $s_{RSC} = \left(\frac{w}{\max w} < 10^{-3}\right)$.

4.1 Synthetic Facial Occlusion Detection

In this section, we use the Extended Yale B database [6] to investigate the performance of SEMC under fixed feature dimension for various synthetic occlusions with varying levels and boundaries. We choose Subsets I and II (717 images, normal-to-moderate lighting conditions) for training and Subset III (453 images, more extreme lighting conditions) for testing. Synthetic occlusions with various boundaries and occlusion levels are imposed on the test samples. The images are resized to 96×84 pixels.

Detection With Various Occlusion Levels. To test the accuracy and stability of SEMC in occlusion detection, we first simulate various levels of occlusions from 10 % to 90 % by replacing a random located block of each test image with a mandrill image. Figure 1a gives nine occluded faces with various occlusion levels and their detailed detection results using the four compared methods. For each method against each occlusion level, the average true positive rates (TPRs) and false positive rates (FPRs) of the occlusion detection results over the 453 test images are also shown in Fig. 1c. For all the cases, the TPRs and FPRs of SEMC are almost always suboptimal compared to the optimal ones of the other methods, whereas the differences between TPRs and FPRs of SEMC are always the largest. This implies that SEMC achieves the optimal balance between TPRs and FPRs. Figure 1c also shows that SEMC achieves its optimal performance at the 50 %˜60 % occlusion levels but not at the lowest ones. The reason is that when the occlusion levels are very low, the dominant differential structures between the mean face (without occlusions) and the occluded face are mainly determined by the differences between faces, which does not be considered by SEMC. The similar problem also exists for the other 3 methods, especially for SSEC.

Different from the other compared methods, SEMC not only detects occlusions but also *understands* occlusion structures. To illustrate this, Fig. 2a counts the number of the structural error metrics selected by SEMC at various mandrill occlusion levels. The main structure of mandrill occlusion actually changes with occlusion levels: when the occlusion levels are low, the texture feature is significant as the edge density is intensive; with the occlusion level increasing, both the texture and color features become more and more significant. This means that for low occlusion levels, SEMC should choose the Laplace-based error metric $\mathcal{E}_{\nabla}$, while for high occlusion levels, SEMC should choose the DE-based error metric $\mathcal{E}_{DE}$. Figure 2a shows that SEMC does perform as expected.

Detection With Various Occlusions. To test the adaptability of SEMC for various occlusions, we simulate 60 % occlusion levels with 7 different objects successively: mandrill, camera, dog, apple, sunflower, random block and white block, i.e., we have 453×7 occluded test images. Figure 1b and d gives the

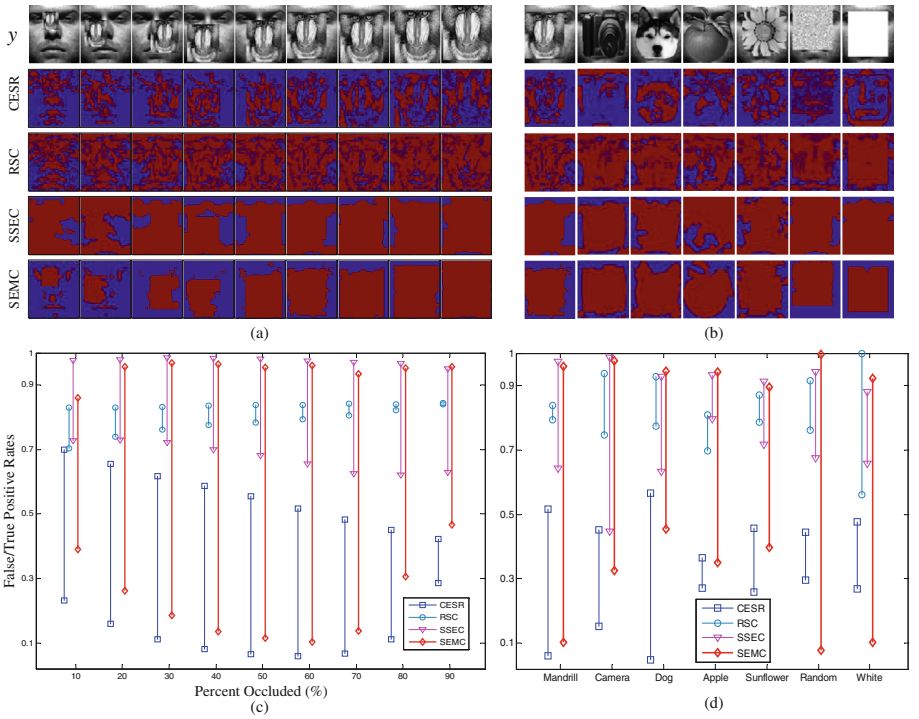

Fig. 1. (a)-(b) The occlusion detection results of various algorithms against various occlusions with various levels on the Extended Yale B database: (a) 10 %˜90 % mandrill occlusions and (b) 60 % various occlusions. (c)-(d) The corresponding average TPRs and FPRs of the 453 occlusion detection results of various algorithms against various occlusions.

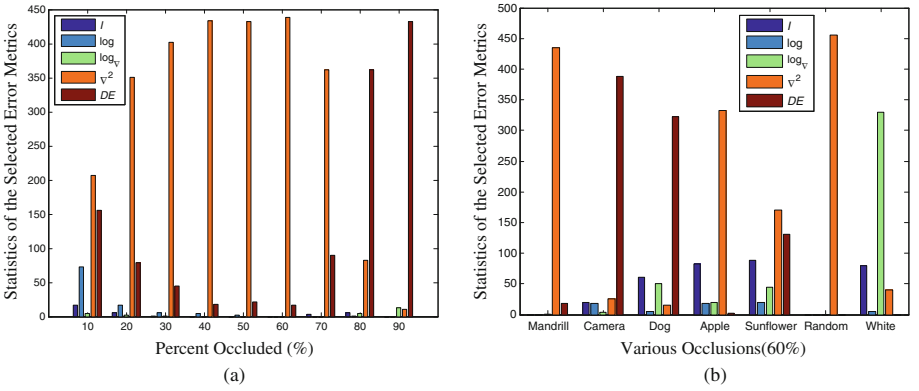

Fig. 2. The statistics of the structural error metrics selected by SEMC for 10 %˜90 % mandrill occlusions (a) and for various occlusions (b).

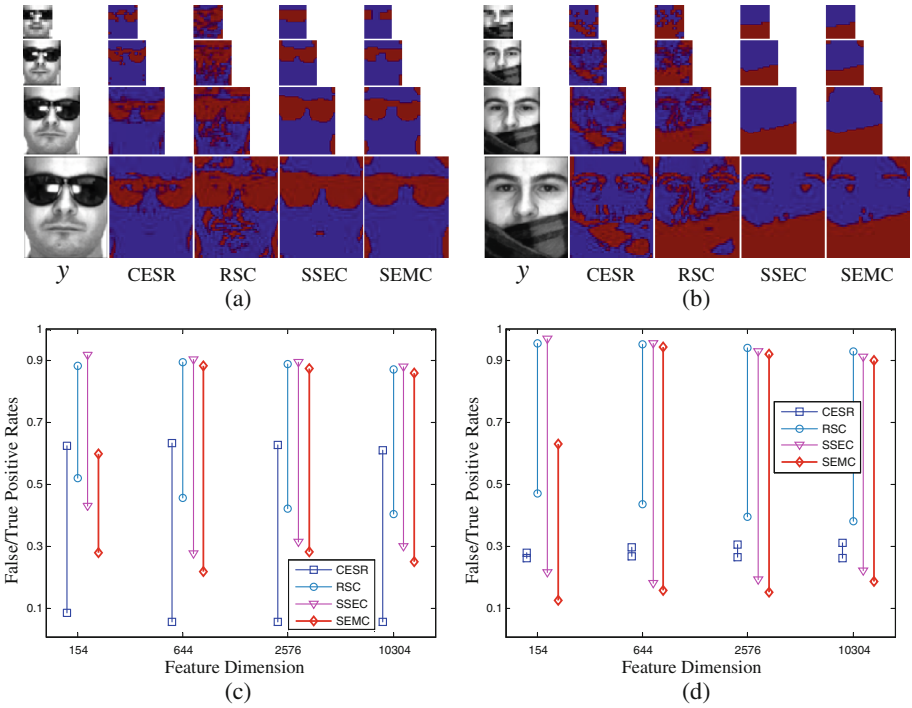

Fig. 3. (a)-(b): The occlusion detection results of various algorithms against various feature dimensions on the AR database. (c)-(d): The average TPRs and FPRs of the occlusion detection results of various algorithms against various feature dimensions on the AR database.

experimental results. Clearly, for all cases, SEMC achieves the optimal balance between TPRs and FPRs. Figure 2b states the occlusion structures that SEMC sees during its detection procedure.

4.2 Real-World Facial Occlusion Detection

We test the performance of SEMC in dealing with real disguises with the AR face database [9]. The grayscale images were resized to 112×92. We select a subset of the database that consists of 119 subjects (65 males and 54 females). For training, we choose 2 unoccluded frontal view images with neutral expressions for each subject from two sessions. For testing, we consider two separate test sets of the 119 subjects. The first/second test set contains 119×2 images of the subjects wearing sunglasses/scarves with neutral expressions from two sessions.

To test the accuracy and robustness of SEMC in occlusion detection for different feature dimensions, we use 4 different downsampled images of dimensions 154, 644, 2576, and 1,0304, which correspond to downsampling ratios of 1/8, 1/4, 1/2, and 1, respectively. The detailed occlusion detection results of the first

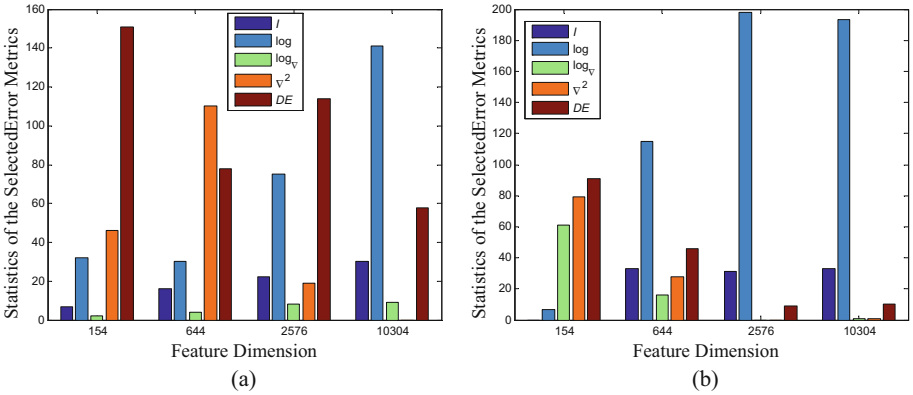

Fig. 4. The statistics of the structural error metrics selected by SEMC for various feature dimensions.

subject in the 4 different dimensions are shown in Fig. 3. Figure 3c and d compare TPRs and FPRs of the detection results of the competing methods. Clearly, SEMC achieves the optimal performance for the two types of disguises except for the lowest feature dimension. The statistical results in Fig. 4 show that with the feature dimension increasing, more and more log-based error metrics are selected by SEMC, since the dark color features of sunglasses/scarves become more and more significant. Figures 3 and 4 demonstrate that occlusion levels affect the performance of SEMC.

5 Conclusions

Most of the state-of-the-art methods in dealing with facial occlusion are based on the alternative iteration of image recovery and occlusion detection. In order to detect facial occlusions efficiently and accurately, we propose a novel method based on the structural error metrics and clustering (SEMC) without image recovery. Experiments show that, even just using the mean face of the training images as the recovery image, SEMC still achieves more accurate and robust performance compared to the related state-of-the-art methods. However, the experiments also show that the minimum occlusion boundary regularity criterion used by SEMC to select the optimal error metric limits its efficacy on occluded images with very low dimension features or with very low occlusion levels. This issue encourages us to further explore new criterion for the optimal error metric selection.

Acknowledgment. This work is partially supported by National Science Foundation of China (61402411, 61379017), Zhejiang Provincial Natural Science Foundation (LY14F020015, LY14F020014), and Program for New Century Excellent Talents in University of China (NCET-12–1087).

References

1. Chen, J., Shan, S., He, C., Zhao, G., Pietikainen, M., Chen, X., Gao, W.: Wld: a robust local image descriptor. IEEE Trans. Pattern Anal. Mach. Intell. **32**(9), 1705–1720 (2010)
2. Deng, W., Hu, J., Guo, J.: Extended src: undersampled face recognition via intraclass variant dictionary. IEEE Trans. Pattern Anal. Mach. Intell. **34**(9), 1864–1870 (2012)
3. Ekenel, H.K., Stiefelhagen, R.: Why is facial occlusion a challenging problem? In: Tistarelli, M., Nixon, M.S. (eds.) ICB 2009. LNCS, vol. 5558, pp. 299–308. Springer, Heidelberg (2009)
4. He, R., Zheng, W., Hu, B.: Maximum correntropy criterion for robust face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1561–1576 (2011)
5. Jia, K., Chan, T.-H., Ma, Y.: Robust and practical face recognition via structured sparsity. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part IV. LNCS, vol. 7575, pp. 331–344. Springer, Heidelberg (2012)
6. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans. Pattern Anal. Mach. Intell. **27**(5), 684–698 (2005)
7. Li, X.X., Dai, D.Q., Zhang, X.F., Ren, C.X.: Structured sparse error coding for face recognition with occlusion. IEEE Trans. Image Process. **22**(5), 1889–1900 (2013)
8. Luan, X., Fang, B., Liu, L., Yang, W., Qian, J.: Extracting sparse error of robust pca for face recognition in the presence of varying illumination and occlusion. Pattern Recogn. **47**(2), 495–508 (2014)
9. Martínez, A.: The ar face database. Technical report, Computer Vision Center (1998)
10. Tuceryan, M., Jain, A.K.: Texture analysis. Handb. Pattern Recogn. Comput. Vis. **276**, 235–276 (1993)
11. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. IEEE Trans. Pattern Anal. Mach. Intell. **34**(12), 2454–2466 (2012)
12. Wandell, B.A.: Foundations of Vision. Sinauer Associates, Sunderland (1995)
13. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2009)
14. Yang, M., Zhang, L.: Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 448–461. Springer, Heidelberg (2010)
15. Yang, M., Zhang, L., Yang, J., Zhang, D.: Robust sparse coding for face recognition. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp. 625–632 (2011)
16. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using markov random fields. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1050–1057 (2009)