Mixed Error Coding for Face Recognition with Mixed Occlusions

Ronghua Liang, Xiao-Xin Li
Zhejiang University of Technology
Hangzhou, China
{rhliang, mordekai}@zjut.edu.cn

Abstract

Mixed occlusions commonly consist in real-world face images and bring with it great challenges for automatic face recognition. The existing methods usually utilize the same reconstruction error to code the occluded test image with respect to the labeled training set and simultaneously to estimate the occlusion/feature support. However, this error coding model might not be applicable for face recognition with mixed occlusions. For mixed occlusions, the error used to code the test image, called the discriminative error, and the error used to estimate the occlusion support, called the structural error, might have totally different behaviors. By combining the two various errors with the occlusion support, we present an extended error coding model, dubbed Mixed Error Coding (MEC). To further enhance discriminability and feature selection ability, we also incorporate into MEC the hidden feature selection technology of the subspace learning methods in the domain of the image gradient orientations. Experiments demonstrate the effectiveness and robustness of the proposed MEC model in dealing with mixed occlusions.

1 Introduction

Real-world face recognition system has to contend with a lot of uncontrolled variations [Hua *et al.*, 2011], such as bad lighting conditions [Lai *et al.*, 2014], large pose variations [Cai *et al.*, 2013], a range of facial expressions [Bettadapura, 2012], apparels, changes in facial hair, eyewear, and partial occlusions. In spite of the diversity, most of these variations can be viewed as different instances of occlusion. When there exists occlusion in a test face image, the classical holistic feature extraction methods, such as Eigenfaces [Turk and Pentland, 1991] and Fisherfaces [Belhumeur *et al.*, 1997], are not applicable since the whole extracted features would be distorted by local occlusion, and robust classifiers, such as SRC [Wright *et al.*, 2009] and CESR [He *et al.*, 2011], might fail to perform recognition due to the common high order statistical structures (localization, orientation, and bandpass) shared by occlusions and face images. Moreover, there often exists practical scenarios that more than one kind of occlusion, which we call *mixed occlusions*, are simultaneously imposed on one face image and greatly increase the difficulty of the recognition task. Apparels and extreme illumination variations are commonly seen mixed occlusions and have long been considered as one of the most difficult problems of face recognition [Tzimiropoulos *et al.*, 2012]. We therefore focus on this challenge in this work.

There are a lot of schemes to deal with occlusion related problems. We put our focus mainly within the framework of sparse coding. In this framework, the test occluded image $y \in \mathbb{R}^m$ is supposed to be a superposition of its clean reconstruction $\hat{y} \in \mathbb{R}^m$ and an error image $e \in \mathbb{R}^m$, and $\hat{y}$ is supposed to be sparsely coded by a linear combination of the training images $A = [A_1, A_2, \cdots, A_K] \in \mathbb{R}^{m \times n}$ of K subjects, where $A_k = \left[a_1^k, a_2^k, \cdots, a_{n_k}^k\right] \in \mathbb{R}^{m \times n_k}$ is a data matrix consisting of n_k training samples from subject k and $n = \sum_{k=1}^{K} n_k$. The error image e is also called the reconstruction error, since it calculates the difference between y and its reconstruction $\hat{y}$. By simultaneously imposing sparse constraint on the coding coefficient and various probabilistic assumptions on the reconstruction error e, researchers have presented a lot of solving methods [Tibshirani, 1996; Chen *et al.*, 2001; Wright *et al.*, 2009; Wright and Ma, 2010; Jia *et al.*, 2012], which can be summarized as

$$\max p(e) \ s.t. \ e = y - Ax, x \text{ is sparse}, \tag{1}$$

where we denote by $p(\cdot)$ a probability density function (PDF) and $x \in \mathbb{R}^n$ is the coding coefficient of y with respect to (w.r.t.) A.

The main limitation of the error coding model (1) is that it usually has a low breakdown point in dealing with the occlusion problems [Wright *et al.*, 2009]. To improve the classification performance of (1), researchers proposed the weighted error coding scheme

$$\max p(e, w) \ s.t. \ e = y - Ax, x \text{ is sparse}, \tag{2}$$

where $w \in \mathbb{R}^m$ is the weight vector. Two representative methods of the weighted error coding model are the Robust Sparse Coding (RSC) [Yang *et al.*, 2011] and the CorrEntropy-based Sparse Representation (CESR) [He *et al.*, 2011]. Due to the capability of feature selection by weighting the features, the weighted error coding model (2) usually outperforms the error coding model (2) in dealing with real-world occlusion. However, the feature selection ability of the

weighted error coding model (2) is still not very strong, since it just weakens (but not removes) the bad features with large errors, and also, due to the shared statistical structures of face images and occlusions, the really bad features not always induce large errors.

To enhance the feature selection ability of the model (2), researchers exploit, instead of the structures of the useful features, the spatial structures shared by occlusions, such as locality, continuity [Lin and Tang, 2007; Zhou *et al.*, 2009] and boundary regularity [Li *et al.*, 2013]. Harnessing the contiguous structure, [Zhou *et al.*, 2009] first built the following error coding model

$$\max p\left(e,s\right)\ s.t.\, e=y-Ax, x \text{ is sparse}, \tag{3}$$

where $s\in\{-1,1\}^m$ denotes the occlusion/feature support, that is, $s_i=-1$ indicates pixel y_i is non-occluded (useful feature) and $s_i=1$ indicates pixel y_i is occluded (useless feature). The two state-of-the-art methods deriving from model (3) are the Sparse Error Correction with Markov Random Fields (SEC_MRF) [Zhou *et al.*, 2009] and the Structured Sparse Error Coding (SSEC) [Li *et al.*, 2013]. Due to efficiently excluding the bad features incurred by occlusion, SEC_MRF and SSEC outperform significantly the methods of model (1) and (2) in dealing with the problem of face recognition with occlusion. However, both SEC_MRF and SSEC require that the number of the training images should be sufficient enough to predict the variations except for occlusions which might consist in the test images. This requirement is hardly available in practice, especially for the scenario of mixed occlusions.

Different from the above three error coding models (1), (2) and (3), [Tzimiropoulos *et al.*, 2012] introduced a robust subspace learning framework in the domain of image gradient orientations (IGO) for appearance-based object recognition. The robustness of the IGO subspace learning methods (such as IGO-PCA and IGO-LDA) owes to the robust IGO features and the capability of hidden feature selection. Specifically, by applying a cosine-based distance measure to the IGO features, the useless features caused by outliers, to some extent, could be automatically eliminated. However, the face recognition experiment in Figure 11 of [Tzimiropoulos *et al.*, 2012] shows that the recognition rates of IGO-PCA drop sharply once the occlusion level exceeds 60%. This indicates that with occlusion level increasing, the IGO-methods might not effectively eliminate the effect of the occlusion.

In this work, we try to solve the problem of face recognition with mixed occlusions by incorporating the robust IGO features into the error coding model (3). For mixed occlusion, we note that the error used to code the test image, we call the discriminative error, and the error used to estimate the occlusion/feature support, we call the structural error, should be coded in different ways. The coding schemes of the two errors are unified in an extended version of the error coding model (3), called the *mixed error coding* (MEC). By considering the MEC model in the IGO domain, we exploit the three conditional PDFs induced by the MEC model, based on which we give the solving algorithm of MEC under the framework of the Gibbs sampling algorithm. Experiments corroborate the effect of the proposed MEC model.

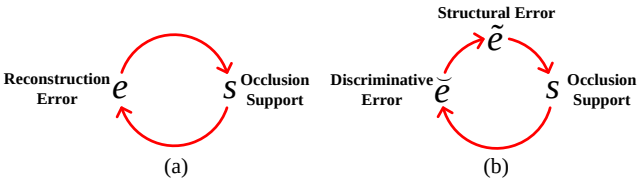

Figure 1: The cyclic interaction relationships between various errors and the occlusion support: (a) the existing model; (b) the proposed mixed error coding model.

2 The Proposed Mixed Error Coding Model

The error model (3) shows that the reconstruction error e actually simultaneously serves for two targets: reconstructing the test image y and estimating the occlusion support s.The interactive relationships of e and s are illustrated in Figure 1 (a). For simple occlusions, the same error might be enough to cope with the two cases, while for mixed occlusions, the scenarios might be completely different. Mixed occlusions usually incur a wide range of variations in the test image and squeeze the useful features into a very low dimensional space, which results in the difficulty of reconstructing y. A common way to solve this problem is to transform the low dimensional features into a high dimensional discriminative feature space (See detail in Subsection 3.1 for constructing high dimensional features in the IGO domain). This will lead to a high dimensional discriminative reconstruction error, which we call the *discriminative error* and denote by $\breve{e}$. Since $\breve{e}$ usually has different dimension and structure with the original test image y, it cannot be directly used to recover the occlusion support s. To estimate s, we need an error that fully considers the spatial and statistical structure of occlusion, which we call the *structural error* and denote by $\tilde{e}$. The structural error $\tilde{e}$ usually has the same dimension with the test image y.

We therefore have three interactive factors: the discriminative error $\breve{e}$, the structural error $\tilde{e}$ and the occlusion support s, whose interactive relationships are illustrated in Figure 1 (b). Clearly, the discriminative error $\breve{e}$ depends on the occlusion support s since s helps to select the discriminative features in calculating $\breve{e}$, and the occlusion support s depends on the structural error $\tilde{e}$ as the spatial and statistical structure of occlusion reflected in $\tilde{e}$ can be used to recover s. We would like to emphasize that the structural error $\tilde{e}$ depends on the discriminative error $\breve{e}$. Actually, the structural error $\tilde{e}$ calculates the structural difference between the test image y and its reconstruction $\hat{y}$, and $\hat{y}$ is reconstructed from the training set A by a sparse coding coefficient x, and finally x is calculated by the discriminative error $\breve{e}$. By formulating the interaction relationships between $\breve{e}$, $\tilde{e}$ and s with a joint PDF $p\left(\breve{e},\tilde{e},s\right)$, we propose a new error coding model, called the *Mixed Error Coding* (MEC), as follows

$$\begin{aligned}\max p\left(\breve{e},\tilde{e},s\right)\ s.t.\ \breve{e}&=\breve{\mathcal{E}}\left(y,Ax\right),\\ \tilde{e}&=\tilde{\mathcal{E}}\left(y,Ax\right), x\geq 0,\end{aligned} \tag{4}$$

where we impose the nonnegative constraint on x to guarantee its sparsity and denote by $\breve{\mathcal{E}}\left(\cdot,\cdot\right)$ and $\tilde{\mathcal{E}}\left(\cdot,\cdot\right)$ the discriminative error metric and the structural error metric, respectively. Since it is difficult to give a definitive formulation of

$p(\breve{e}, \tilde{e}, s)$, according to the interaction relationships between $\breve{e}$, $\tilde{e}$ and s shown in Figure 1 (b), we adopt the Gibbs sampling algorithm to solving (4) as follows

$$\left(\breve{e}^{(t)}, x^{(t)}\right) = \arg\max_{\breve{e},x} p\left(\breve{e}\left|s^{(t-1)}\right.\right), \quad s.t.\ \breve{e} = \breve{\mathcal{E}}\left(y, Ax\right), x \geq 0, \tag{5}$$

$$\tilde{e}^{(t)} = \arg\max_{\tilde{e}} p\left(\tilde{e}\left|\breve{e}^{(t)}\right.\right), \quad s.t.\ \tilde{e} = \tilde{\mathcal{E}}\left(y, Ax^{(t)}\right), \tag{6}$$

$$s^{(t)} = \arg\max_{s} p\left(s\left|\tilde{e}^{(t)}\right.\right), \tag{7}$$

where the superscript (t) denotes the tth iteration.

3 Three Statistical Inferences for Mixed Error Coding in IGO Domain

To infer the three interacted factors $(\breve{e}, \tilde{e}, s)$ from (5), (6) and (7), a primary task is first to establish the three conditional probabilistic density/mass functions $p(\breve{e}|s)$, $p(\tilde{e}|\breve{e})$ and $p(s|\tilde{e})$. Due to the robustness of the IGO features we discuss in the introduction, we explore each of the three conditional probabilistic models in the IGO domain in the following subsections. For convenience, we denote by $\Phi(\cdot)$ the IGO transformation function and briefly denote the IGO features of y and A by $\mathring{y} = \Phi(y)$ and $\mathring{A} = \Phi(A)$, respectively.

3.1 Gaussian Distribution of Discriminative Error

We first consider the conditional PDF $p(\breve{e}|s)$. As an important discriminative information to search for the right subspace (class) that the test image comes from, the discriminative error $\breve{e}$ should be only considered in the non-occluded region to avoid the effect incurred by occlusion. To compensate for the losses caused by occlusion, we transform the features in the non-occluded region into a high dimensional feature space. As demonstrated by [Wright *et al.*, 2009], in the framework of sparse coding, it is the dimension of the feature that plays a critical role in determining the discriminability. For the IGO features, [Tzimiropoulos *et al.*, 2011] suggested imposing cosine and sine kernel on the angular data and then stacking the outputs as columns to form new feature vectors, that is, the new high dimensional features of the test and training images can be formulated as $\breve{y} = \left[\cos\mathring{y}^T \ \sin\mathring{y}^T\right]^T \in \mathbb{R}^{2m}$ and $\breve{A} = \left[\cos\mathring{A}^T \ \sin\mathring{A}^T\right]^T \in \mathbb{R}^{2m\times n}$. Then, in this high dimensional feature space, it is reasonable to assume that the DE $\breve{e}$ follows the Gaussian distribution with zero mean

$$p(\breve{e}|s) = \prod_{i=1}^{2m} \frac{1}{\sqrt{2\pi}\breve{\sigma}} \exp\left(-\frac{(1-\breve{s}_i)\,\breve{e}_i^2}{2\breve{\sigma}^2}\right), \tag{8}$$

where $\breve{e} = \breve{y} - \breve{A}x$, $\breve{s} = \left[s^T \ s^T\right]^T$ and $\breve{\sigma}$ is the Gaussian kernel size. Figure 3 (b) shows the histograms of 5 discriminative errors produced during the iteration of our proposed algorithm for a test image.

For inference of $\breve{e}$, substituting (8) into (5), we have

$$\left(\breve{e}^{(t)}, x^{(t)}\right) = \arg\min_{\breve{e},x} \left\|\frac{1}{2}\left(1-\breve{s}^{(t-1)}\right)\odot\breve{e}\right\|_2^2 \quad s.t.\ \breve{e} = \breve{y} - \breve{A}x, x \geq 0, \tag{9}$$

where $\odot$ is the Hadamard product. The optimization problem (9) is a nonnegative least squares problem (NNLS) and can be solved by the classical active set algorithm [He *et al.*, 2011].

3.2 Uniform Distribution of Structural Error

We now consider the conditional PDF $p(\tilde{e}|\breve{e})$. The structural error $\tilde{e}$ reflects the structural difference between the test image $\mathring{y}$ and its reconstruction $\hat{\mathring{y}}$. Using the coding coefficient $x^{(t)}$ solved from (9), we could reconstruct a clean version $\hat{\mathring{y}}^{(t)} = \Phi\left(Ax^{(t)}\right)$ of the test image $\mathring{y}$. In the original pixel domain, [Li *et al.*, 2013] pointed out that the structural error between y and its reconstruction $\hat{y}$ calculated by a well designed error metric would present a special distribution structure: the errors corresponding to the occluded part concentrate on a large value, and the errors corresponding to the non-occluded part concentrate on zero. However, in the IGO domain, we will show that the structural error measured by a well designed structural error metric would follow a uniform distribution which is independent of the occlusion support. Therefore, what is critical for building the conditional PDF $p(\tilde{e}|\breve{e})$ is to design a reasonable structural error metric.

We first review a robust distance measure in the IGO domain. [Tzimiropoulos *et al.*, 2012] statistically verified that the gradient orientation differences of any two pixel-wise dissimilar images follow a uniform distribution in the interval $[-\pi, \pi)$ with a high significant level. Furthermore, this uniform distribution is subtly used to approximately cancel out the error caused by outliers according to the following theorem:

Theorem 1. *Let $u(\cdot)$ be a mean ergodic stochastic process and $u(t)$ follows a uniform distribution in $[-\pi, \pi)$, then for any non-empty interval $\mathcal{X} \in \mathbb{R}$, $\int_{\mathcal{X}} \cos(u(t))\,dt = 0$.*

Based on Theorem 1, [Tzimiropoulos *et al.*, 2012] defined a cosine-based distance measure (CDM):

$$\mathrm{CDM}\left(\mathring{y}, \hat{\mathring{y}}\right) \triangleq \sum_{i=1}^{m}\left(1-\cos\left(\mathring{y}_i - \hat{\mathring{y}}_i\right)\right). \tag{10}$$

Ideally, according to Theorem 1, all of the errors incurred by outliers would be summed up to zero in (10) and hence CDM only calculates the distance of the clean parts of the two compared images.

In spite of its robustness, CDM cannot be directly used to estimate the distribution of the error between two compared images, since its calculating result is a scalar. We therefore need to vectorize CDM to form a new error metric. In fact, if we shrink the action scope of the CDM to the neighborhood of one feature point of the two compared images and use the distance calculated by CDM on this neighborhood as the final observed error value at this point, we then obtain the following cosine-based error metric (CEM)

$$\mathrm{CEM}\left(\mathring{y}_i, \hat{\mathring{y}}_i\right) \triangleq \sum_{j\in\mathcal{N}(i)}\left(1-\cos\left(\mathring{y}_j - \hat{\mathring{y}}_j\right)\right), \tag{11}$$

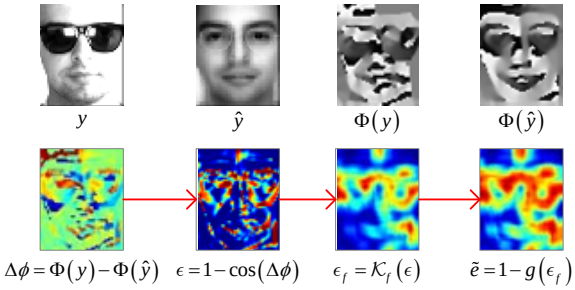

Figure 2: The calculating flow of the proposed structural error metric $\tilde{\mathcal{E}}(\cdot,\cdot)$ in (13). Here, $\mathcal{K}_f(\cdot)$ is a K-means filtering operator and $g(\cdot)$ is a Gaussian kernel.

where $\mathcal{N}(i)$ is the neighborhood of feature point i. A fast implementation of CEM can be given by imposing K-means filtering on the error vector $\epsilon = 1 - \cos\left(\mathring{y} - \mathring{\hat{y}}\right)$. Note that CEM only harnesses the spatial (local and contiguous) structure of occlusion. In order to enhance the clustering effect of CEM, we further consider the statistical structure of the error calculated by CEM. To integrate the statistical local information of the input error, [He *et al.*, 2011] suggested using the correntropy induced metric (CIM)

$$\mathrm{CIM}(e_i) \triangleq 1 - \exp\left(\frac{e_i^2}{2\sigma^2}\right). \tag{12}$$

We therefore incorporate CEM into CIM to form a new structural error metric

$$\tilde{\mathcal{E}}\left(\mathring{y}_i, \mathring{\hat{y}}_i\right) \triangleq \mathrm{CIM}\left(\mathrm{CEM}\left(\mathring{y}_i, \mathring{\hat{y}}_i\right)\right). \tag{13}$$

Figure 2 illustrates the whole calculating flow of (13). In essence, the structural error metric (13) amounts to orderly imposing three smoothing operations (cosine, K-means and Gaussian) on the gradient orientation difference $\mathring{y} - \mathring{\hat{y}}$, which finally smooth the diversity of the whole error values and thus induce a uniform distribution of the measured SE $\tilde{e}$ on the interval $[0, 1)$. The experiment in Figure 3 (f) demonstrates that $p(\tilde{e}|\breve{e})$ is more and more close to a uniform distribution with the improvement of the quality of the recovering image $\hat{y}$.

The uniform distribution of the structural error $\tilde{e}$ calculated by (13) can not be used to infer $\tilde{e}$, but can be utilized as a stopping iterative criterion of our proposed algorithm, as shown in Figure 3 (f). We calculate $\tilde{e}$ by directly inputting the entries of $\mathring{y}$ and $\mathring{\hat{y}}^{(t)} = \Phi\left(Ax^{(t)}\right)$ into the error metric (13).

3.3 Bayesian Inference of Occlusion Support

We now consider the conditional probabilistic mass function (PMF) $p(s|\tilde{e})$. Since s and $\tilde{e}$ have the same dimension and similar structure (i.e., the larger the structural error $\tilde{e}_i$, the more tendency $s_i = 1$), we explore $p(s|\tilde{e})$ from a *Bayesian* perspective

$$p(s|\tilde{e}) \propto p(\tilde{e}|s)\, p(s), \tag{14}$$

where $p(s|\tilde{e})$ is viewed as a posterior PMF, $p(\tilde{e}|s)$ is the likelihood function, and $p(s)$ is the prior probability of s. This Bayesian model has been explored deeply in [Zhou *et al.*, 2009]. By summarizing the work of [Zhou *et al.*, 2009] and incorporating (14) into (7), we have the following optimization problem

$$s^{(t)} = \arg\max_{s} \sum_{i=1}^{m} \sum_{j \in \mathcal{N}(i)} \lambda_s s_i s_j + \sum_{i=1}^{m} \left(1 - \frac{3}{2}\mathcal{K}_\tau\left(\tilde{e}_i^{(t)}\right) + \lambda_\mu\right) s_i \tag{15}$$

where $\mathcal{K}_\tau(\tilde{e}_i) = \begin{cases} 1, & |\tilde{e}_i| > \tau \\ 0, & |\tilde{e}_i| \le \tau \end{cases}$. (15) can be solved using graph cuts [Kolmogorov and Zabih, 2004].

4 The Algorithm

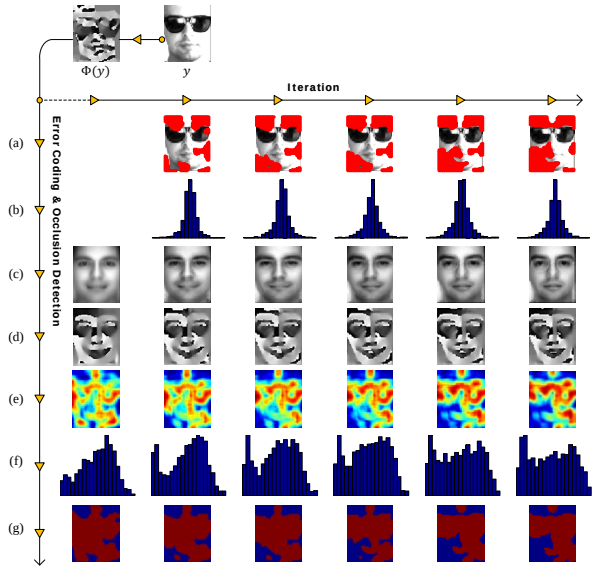

Figure 3: Detailed illustration of the iterative procedure of the proposed algorithm. (a) The selected features marked by red point. (b) The histograms (Gaussian distribution) of the discriminative errors. (c) The reconstructed images. (d) The IGO transformed faces of the reconstructed images. (e) The structural errors. (f) The histograms (uniform distribution) of the structural errors. (g) The detected occluded points marked by red points.

By iteratively using (9), (13) and (15), we could solve the MEC problem (4). The important outputs of this iterative procedure are the the occlusion/feature support s and the coding coefficient x. Using s and x, we are able to identify the test image y from the training set A based on some measure of goodness-of-fit. In this work, we adopt a subject specific reconstruction classifier similar to the sparse classifier proposed

by [Wright *et al.*, 2009], but the major difference is that the classifier introduced here is based on the selected features.

Algorithm 1 summarizes the whole procedure of optimizing the MEC model (4). Figure 3 gives a detailed illustration of the iterative procedure.

Algorithm 1 Mixed Error Coding (MEC)

Input: training data A, test sample y.
Output: occlusion/feature support s, identity (y).

1. Calculate the mean face $\hat{y}$ of the training image A;
2. Transform y, $\hat{y}$ and A into the IGO domain: $\mathring{y} = \Phi(y)$, $\mathring{\hat{y}} = \Phi(\hat{y})$, $\mathring{A} = \Phi(A)$;
3. **Repeat**
4. Calculate the structural error $\tilde{e}$: $\forall i \in \{1, 2, \cdots, m\}$, $\tilde{e}_i = \mathrm{CC}\left(\mathring{y}_i, \mathring{\hat{y}}_i\right)$;
5. Recover the occlusion support s by solving (15);
6. Calculate the discriminative error $\breve{e}$ and the coding coefficient x by solving (9);
7. Calculate the reconstruction $\mathring{\hat{y}} = \Phi(Ax)$;
8. **Until** maximum iterations or convergence.
9. For $k = 1, \cdots, K$, compute the residuals $r_k = \left\| \frac{1}{2}(1 - \breve{s}) \odot \left(\breve{y} - \breve{A}\delta_k(x)\right) \right\|_2^2$, where $\delta_k(x)$ is a new vector whose only nonzero entries are the ones in x that correspond to subject k;
10. identity $(y) = \arg\min_k r_k$.

5 Experiments

To evaluate the performance of the proposed MEC algorithm, we compare it with four related popular methods for robust face recognition with occlusion: IGO-PCA [Tzimiropoulos *et al.*, 2012], SSEC [Li *et al.*, 2013], CESR [He *et al.*, 2011] and RSC [Yang *et al.*, 2011]. The parameters of our MEC algorithm are selected as: $\lambda_\mu = 0$, $\tau = 0.3$,$\lambda_s = 2$ and $\sigma = 0.75$ in CIM. The parameters of the other methods are set according to the strategy suggested in their papers.

We conduct a set of experiments on the Extended Yale B database [Georghiades *et al.*, 2001] and the AR database [Martínez, 1998]. The Extended Yale B database contains 2414 frontal face images of 38 persons under 64 different illumination conditions. The AR database, which is one of the very few databases that contain real disguise, consists of over 4,000 color images corresponding to 126 persons' frontal view faces with different facial expressions (neutral, smile, anger, and scream), illumination conditions (left light on, right light on, and all side lights on), and occlusions (sunglasses and scarves). 26 pictures were taken for each person in two sessions.

5.1 Synthetic Occlusions Mixed with Various Illuminations

In this experiment, we use the Extended Yale B database to test the robustness of our algorithm against various levels of synthetic occlusions under various illumination conditions. The illumination conditions of the Extended Yale B database are partitioned to 5 subsets: from normal illumination variations to extreme ones. For training, we use images from Subset I and II (717 images, with normal-to-moderate illumination conditions); for testing, we use images from Subset III (453 images, with extreme illumination conditions), Subset IV (524 images, with more extreme illumination conditions) and Subset V (712 images, with the most extreme illumination conditions), respectively. To simulate different levels (from 0% to 90%) of contiguous occlusion, we replace a random block of each test image with a mandrill image, which has similar structure with the human face and has been widely used as synthetic occlusion in robust face recognition testing [Wright *et al.*, 2009; Zhou *et al.*, 2009; Yang *et al.*, 2011; He *et al.*, 2011; Li *et al.*, 2013].All images are cropped and resized to 96 × 84 pixels. Note that this experimental setting is similar to the one in [Wei *et al.*, 2012].

Figure 4 compares the recognition rates of MEC with other related approaches on the three different test subsets, respectively. Compared to MEC and IGO-PCA, the recognition performances of SSEC, CESR and RSC drop sharply with the illumination conditions worsening from Subset III to IV, which manifests the importance of the IGO features against illumination changes. Clearly, in this experiment, the nearest competitor of MEC is IGO-PCA, which performs almost as well as MEC when the occlusion level is lower than some breakdown point. This breakdown point, however, rapidly declines with the illumination conditions of the test images worsening. This illustrates that the illumination conditions greatly intensify the effect of occlusions in face recognition. Especifically for the most extreme illumination conditions of Subset V, the breakdown point almost decreases to 0 and MEC significantly outperforms IGO-PCA, as seen from Figure 4 (c). This indicates the important role of the feature selection of MEC.

5.2 Real-world Disguises Mixed with Highlight Illuminations

We next test our algorithm on real disguises mixed with nonuniform illuminations using the AR Face database. We select a subset of the database that consists of 65 male subjects and 54 female subjects. The grayscale images are resized to resolution 112 × 92. For training, we use 952 non-occluded frontal view images (8 images for each subject) with varying facial expressions but normal illuminations. For testing, we use images that simultaneously contain illumination variations (normal/right-side/left-side light conditions) and disguises (sunglasses and scarves).

Figure 5 compares the recognition rates of different methods using different downsampled images of dimensions 154, 644, 2576, and 1,0304, which correspond to downsampling

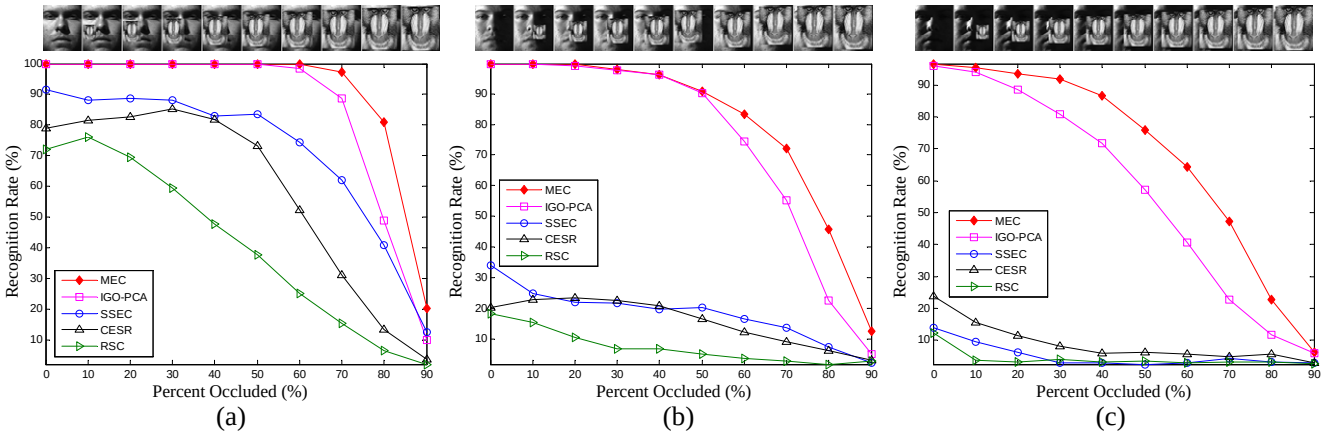

Figure 4: Recognition against mandrill occlusion with various levels (0%~90%) on the Extended Yale B database. The synthetic occluded images are imposed on 3 different illuminations subsets, respectively: (a) Subset III, (b) Subset IV, and (c) Subset V.

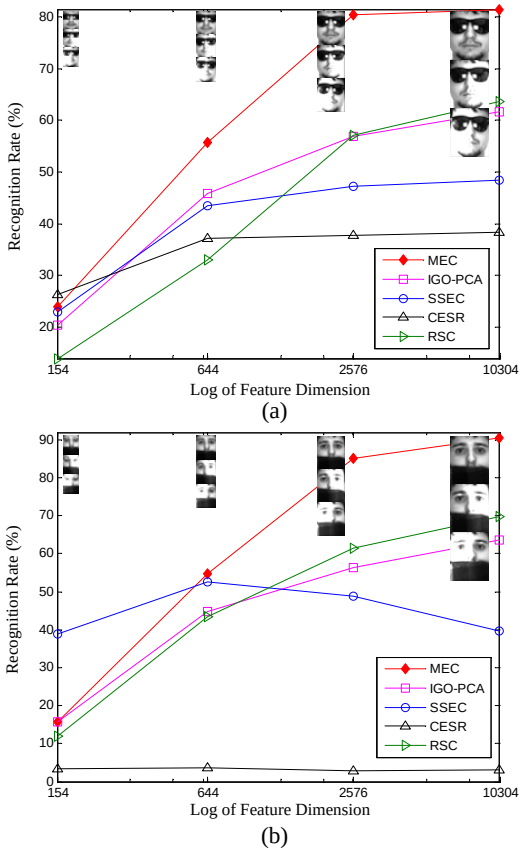

Figure 5: Recognition rates of various algorithms under various feature spaces against real-world occlusions in the AR database: (a) sunglasses occlusion, (b) scarf occlusion.

ratios of 1/8, 1/4, 1/2, and 1, respectively. With feature dimensions increasing, the recognition performance of our MEC dramatically outperforms the other four methods. Specially, at the dimension of 1,0304, compared to the nearest competitors IGO-PCA and RSC, MEC achieves over 20% and 25% higher recognition rates for the sunglasses and scarf disguises, respectively. Note that the recognition rate gaps between MEC and IGO-PCA in Figure 4 are much smaller than the ones in Figure 5. This is mainly caused by the opposite illumination conditions: the illuminations in Figure 4 are dark and shadowed, while the illuminations in Figure 5 are mostly with highlight conditions. Clearly, the high specularities cannot be well normalized in the IGO domain. In this scenario, feature selection becomes very important. The significantly higher recognition rates of our MEC than those of the other four methods at high dimensions show that MEC has a strong feature selection ability in high feature space. We also observe that the recognition rate of SSEC against scarf disguises (see Figure 5 (b)) significantly outperforms the other compared methods at the low dimension. This inspires us to further study the occlusion structure of the low dimensional image from SSEC in the IGO domain in the future.

6 Conclusions

We propose a mixed error coding (MEC) model for robust classification of face images with mixed occlusions. To cope with the complex variations caused by mixed occlusions, MEC distinguishes the error used to code the occluded image from the error used to estimate its occlusion/feature support, and explores the statistical correlations between the two various errors and the occlusion support in the domain of image gradient orientations. An important result of this exploration is a well-designed structural error metric, which fully uses the spatial and statistical structure of the gradient orientation differences of two compared images and plays an important role in feature selection. Experiments demonstrate the effectiveness and robustness of MEC in dealing with mixed occlusions.

Acknowledgment

This work is partially supported by National Science Foundation of China (61402411), Zhejiang Provincial Natural Science Foundation (LY14F020015), and Program for New Cen-

tury Excellent Talents in University of China (NCET-12-1087).

Biography

Ronghua Liang received the Ph.D. in computer science from Zhejiang University in 2003. He is currently a Professor of Computer Science and Vice Dean of College of Information Engineering, Zhejiang University of Technology, China. He has published more than 50 papers in leading international journals and conferences including IEEE TKDE, IEEE TVCG, IEEE VIS. His research interests include Information Visualization, Computer Vision, and Medical Visualization.

Xiao-Xin Li is currently an Assistant Professor with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou, China. His research interests include Image Processing and Machine Learning.

References

[Belhumeur *et al.*, 1997] P Belhumeur, J Hespanha, and D Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[Bettadapura, 2012] Vinay Bettadapura. Face expression recognition and analysis: the state of the art. *arXiv preprint arXiv:1203.6722*, 2012.

[Cai *et al.*, 2013] Xinyuan Cai, Chunheng Wang, Baihua Xiao, Xue Chen, and Ji Zhou. Regularized latent least square regression for cross pose face recognition. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, pages 1247–1253. AAAI Press, 2013.

[Chen *et al.*, 2001] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM review*, pages 129–159, 2001.

[Georghiades *et al.*, 2001] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.

[He *et al.*, 2011] R. He, W. Zheng, and B. Hu. Maximum correntropy criterion for robust face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1561–1576, 2011.

[Hua *et al.*, 2011] Gang Hua, Ming-Hsuan Yang, Erik Learned-Miller, Yi Ma, Matthew Turk, David J. Kriegman, and Thomas S. Huang. Introduction to the special section on real-world face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):1921–1924, 2011.

[Jia *et al.*, 2012] Kui Jia, Tsung-Han Chan, and Yi Ma. Robust and practical face recognition via structured sparsity. In *European Conference on Computer Vision*, pages 331–344. Springer, 2012.

[Kolmogorov and Zabih, 2004] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, 2004.

[Lai *et al.*, 2014] Z. Lai, D. Dai, C. Ren, and K. Huang. Multilayer surface albedo for face recognition with reference images in bad lighting conditions. *IEEE Transactions on Image Processing*, 23(11):4709–4723, 2014.

[Li *et al.*, 2013] Xiao-Xin Li, Dao-Qing Dai, Xiao-Fei Zhang, and Chuan-Xian Ren. Structured sparse error coding for face recognition with occlusion. *IEEE Transactions on Image Processing*, 22(5):1889–1900, 2013.

[Lin and Tang, 2007] Dahua Lin and Xiaoou Tang. Quality-driven face occlusion detection and recovery. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 1–7, 2007.

[Martínez, 1998] A.M. Martínez. The ar face database. Technical report,, Computer Vision Center, 1998.

[Tibshirani, 1996] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288, 1996.

[Turk and Pentland, 1991] M Turk and A Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[Tzimiropoulos *et al.*, 2011] Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. Sparse representations of image gradient orientations for visual recognition and tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 26–33, Colorado Springs, USA, 2011. IEEE.

[Tzimiropoulos *et al.*, 2012] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. Subspace learning from image gradient orientations. *IEEE transactions on pattern analysis and machine intelligence*, 34(12):2454–66, 2012.

[Wei *et al.*, 2012] Xingjie Wei, Chang-Tsun Li, and Yongjian Hu. Robust face recognition under varying illumination and occlusion considering structured sparsity. In *International Conference on Digital Image Computing Techniques and Applications*, pages 1–7, 2012.

[Wright and Ma, 2010] J. Wright and Yi Ma. Dense error correction via ℓ^1-minimization. *IEEE Transactions on Information Theory*, 56(7):3540–3560, 2010.

[Wright *et al.*, 2009] J Wright, AY Yang, A Ganesh, SS Sastry, and Y Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210–227, 2009.

[Yang *et al.*, 2011] Meng Yang, Lei Zhang, Jian Yang, and David Zhang. Robust sparse coding for face recognition. In *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pages 625–632, 2011.

[Zhou *et al.*, 2009] Zihan Zhou, A. Wagner, H. Mobahi, J. Wright, and Yi Ma. Face recognition with contiguous occlusion using markov random fields. In *Proc. IEEE Int'l Conf. Computer Vision*, pages 1050–1057, 2009.