Structured Sparse Error Coding for Face Recognition With Occlusion

Xiao-Xin Li, Dao-Qing Dai, *Member, IEEE*, Xiao-Fei Zhang, and Chuan-Xian Ren

***Abstract*— Face recognition with occlusion is common in the real world. Inspired by the works of structured sparse representation, we try to explore the structure of the error incurred by occlusion from two aspects: the error morphology and the error distribution. Since human beings recognize the occlusion mainly according to its region shape or profile without knowing accurately what the occlusion is, we argue that the shape of the occlusion is also an important feature. We propose a morphological graph model to describe the morphological structure of the error. Due to the uncertainty of the occlusion, the distribution of the error incurred by occlusion is also uncertain. However, we observe that the unoccluded part and the occluded part of the error measured by the correntropy induced metric follow the exponential distribution, respectively. Incorporating the two aspects of the error structure, we propose the structured sparse error coding for face recognition with occlusion. Our extensive experiments demonstrate that the proposed method is more stable and has higher breakdown point in dealing with the occlusion problems in face recognition as compared to the related state-of-the-art methods, especially for the extreme situation, such as the high level occlusion and the low feature dimension.**

***Index Terms*— Face recognition, high-breakdown point classification, malicious occlusion, outlier detection, structured sparse representation.**

I. Introduction

WHILE image occlusion is an important cue for scene understanding [1] and video segmentation [2], it poses a significant obstacle for robust, real-world face recognition (FR). Since the work of Martinez [3], a variety of robust methods have been proposed for matching partially occluded test images to unoccluded training samples [4]–[12]. For almost all of these methods, there exists a breakdown point[1] beyond which these methods are very sensitive to occlusions and their FR performances drop *sharply*. Why is high-level facial occlusion a challenging problem? Experiments in [8], [10], [12], and [13] illustrate that missing discriminative information caused by occlusion only account for a very small part of the performance drop. Wright *et al*. [14] recently show that accurate recovery of sparse signals is possible and computationally feasible even with nearly 100% of the observations corrupted. Hence, the question of questions is how to recognize the occluded regions or pixels accurately.

If the error $e \in \mathbb{R}^m$ between the occluded face image $y \in \mathbb{R}^m$ and its ground truth $y^0 \in \mathbb{R}^m$ is known, i.e., $e = y - y^0$, it will be easy to recognize the occluded regions, and the error support $s \in \{-1, 1\}^m$ [9], where $s_i = -1$ if $e_i = 0$ and $s_i = 1$ if $e_i \neq 0$[2], indicates the occluded pixels of y. In FR, however, y^0 is usually unknown, and an approximation of y^0 is usually calculated by projecting y onto the subspace Ω_A spanned by the training samples $A = [A_1, A_2, \ldots, A_K] \in \mathbb{R}^{m \times n}$ of K subjects, where $A_k = \left[a_1^k, a_2^k, \ldots, a_{n_k}^k\right] \in \mathbb{R}^{m \times n_k}$ is a data matrix consisting of n_k training samples from the kth class and $n = \sum_{k=1}^{K} n_k$. Suppose that y belongs to one of the K subjects[3], and $x \in \mathbb{R}^n$ is the projected coefficient of y in Ω_A, the projection error

$$\hat{e} = y \ominus \hat{y} = y \ominus Ax, \tag{1}$$

where $\ominus$ denotes any error metric operator, is usually used to estimate the probability with which the image pixel y_i is occluded [10], [11] or the likelihood between the error e and the error support s [9]. The error coding model (1) shows that the error $\hat{e}$ is determined by three interacted factors: the training and test samples A and y, the coding scheme of y with respect to (w.r.t.) A, and the error metric $\ominus$. In real-world FR, variations among the training and test samples themselves contribute enormously to the errors. Except for the occlusion/corruption, the changes in illumination/pose/expression [3], [15], low-resolution [16], image misalignment [13], etc., all account for the errors. To simplify matters, we just focus on the facial occlusion problem, and assume that any variation, except occlusion, emerging in the test image, simultaneously emerges in part of the training ones, i.e., the unoccluded regions of the test image can be well approximated by the corresponding regions of the

Manuscript received May 9, 2012; revised December 3, 2012; accepted December 18, 2012. Date of publication January 4, 2013; date of current version March 14, 2013. This work was supported in part by the National Science Foundation of China, under Grant 11171354, Grant 90920007, and Grant 61203248, the Ministry of Education of China, under Grant SRFDP-20120171120007 and Grant SRFDP-20120171110016, and the Post-Doctoral Science Foundation of China, under Grant 2011M501361. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Alex C. Kot.

The authors are with the Center for Computer Vision and Department of Mathematics, Sun Yat-Sen University, Guangzhou 510275, China (e-mail: mordekai@qq.com; stsddq@mail.sysu.edu.cn; zhangxf9@mail2.sysu.edu.cn; pr.renchx@gmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2013.2237920

[1] In this work, the breakdown point represents the lowest occlusion level or feature dimension of an image that can cause an estimator to produce arbitrarily bad results.

[2] We use z_i to denote the ith entry of the vector z.

[3] Due to occlusions, the existing techniques, such as sparsity concentration index (SCI) [8], are not suitable to make a validation test. The validation test method for FR with occlusion will be our future work. In this work, we just suppose y belongs to one of the training subjects.

training ones. Based on this assumption, we concentrate on the coding scheme of y w.r.t. A, and the error metric $\ominus$ in the following.

The success of manifold learning implies that the high dimensional face images can be sparsely represented or coded by the representative samples on the manifold, which leads most of the state-of-the-art FR techniques code y w.r.t. A using sparse representation (SR) [8]–[11], and [17]. Meanwhile, the subtractive operator $SUB\left(y_i, \hat{y}_i\right) = \left|y_i - \hat{y}_i\right|$ is commonly used as the error metric. Based on SR and the SUB error metric, the error model has been explored extensively under different priori distribution assumptions of the error $\hat{e}$. The assumption that $\hat{e}$ follows the Gaussian distribution leads to the ℓ_2/ℓ_1 model (LASSO [18]): $\min_x \|y - Ax\|_2^2 \; s.t. \; \|x\|_1 \leq \epsilon$, where $\epsilon > 0$ is a constant; the assumption that $\hat{e}$ follows the Laplacian distribution leads to the ℓ_1/ℓ_1 model [8], [14]: $\min_x \|y - Ax\|_1 \; s.t. \; \|x\|_1 \leq \epsilon$. Recently, Yang *et al.* [11], [19] argue that the error $\hat{e}$ "may be far from Gaussian or Laplacian distribution, especially when there are occlusions, corruptions and/or other variations", and propose a general probabilistic model assumption, called robust sparse coding (RSC). Note that all of the above methods measure the error $\hat{e}$ by the traditional SUB operator. The error metric $\ominus$ does not attract attentions until the work of He *et al.* [10]. Using a robust error metric, called the correntropy induced metric (CIM) [20], He *et al.* [10] propose the correntropy-based SR (CESR) scheme. Interestingly, although CESR and RSC are based on different theory fundamentals, they are both in spirit of the weighted sparse coding (WSC) model: $\min \|w \odot (y - Ax)\|_2^2 + \phi(w) \; s.t. \; x$ is sparse,where $\odot$ is the Hadamard product (i.e., the element wise multiplication of two vectors), $w \in \mathbb{R}^m$ is a weighted vector, $\phi(w)$ is a cost function of w. This means there must exist some close relationship between the error metric and error distribution.

Although all of the above SR based methods aim at dealing with both corruption and occlusion, the experiments show that they can cope with corruption much better than occlusion. The main reason might be that the randomly corrupted pixels have strikingly diverse structures with the real face images, while the contiguous occlusions, in general, consisting of natural images, share high order statistical structures (localization, orientation, and bandpass) [21] with face images. These shared statistical structures might reduce the discriminability of the SR based methods. In addition, we discover that the features in the occluded regions of the test image tend to determine the recognition results of the SR based methods, if features similar to occlusion also exist in the corresponding regions of the training samples. We call these features *Locally Salient Features* (LSF) incurred by occlusions (see Section A of the supplementary appendix for more details). Clearly, with occlusion level increasing, such LSFs will play more and more important roles. In order to improve the FR performance for high level occlusion, Zhou *et al.* [9] incorporate the continuity of the error support into the error structure and propose a robust algorithm, called Sparse Error Correction with Markov Random Fields (SEC_MRF). Although SEC_MRF greatly improves the recognition performance, the FR experiment for the synthetic occlusion with the Extended Yale B in [9] shows that the FR rate falls sharply from 88.5% to 40.3% when the occlusion level increases from 80% to 90%. Such a big drop shows that a more intrinsic error structure, besides sparsity and continuity, should be explored for describing the contiguous occlusion.

In this work, we study the intrinsic error structure of contiguous occlusion from two sides: *i*) the structure of the error support, including its spatial continuity and its shape; *ii*) the structure of the error distribution under different error metrics. The study of the error structure makes our work contribute mainly in the following aspects:

1) A *morphological graph* model for the error support structure. Recently, structured sparse representation (SSR) has attracted a lot of attention [22]–[25], since the measurements used to recover the structured sparse signal can be substantially decreased compared to the standard compressive sensing method [26]. Using the graph theory, SSR also offers a scheme to use the coefficient structure [25]. Inspired by these works, in order to explore the geometrical structure of the occlusion error, we propose a new graph model, called *morphological graph*, by integrating the morphology information of occlusion into graph.
2) An exponential probabilistic model for the error distribution structure. Although there must exist some close relationship between the error metric and error distribution, to the best of our knowledge, none of the existing methods takes the two factors together. We study the error distribution under different error metrics, especially, under the entropy-induced metrics [20]. With the assumption that the unoccluded region of the test image y can be approximated sufficiently by the corresponding regions of the training samples, we show that the error $\hat{e}$ measured by the combination operator of the correntropy induced metric and the dilation invariance metric, might present a special sparse structure: both the error in the unoccluded region and the error in the occluded region follow the exponential distribution, which induces good *source-and-error separation*.
3) Using the structured information of the error support and error distribution, we propose a new scheme, called *Structured Sparse Error Coding* (SSEC), for FR with occlusion. Experiments show that SSEC is a high breakdown point method which achieves much stable recognition performance for various occlusion levels and various feature dimensions.

The rest of this paper is organized as follows: Section II explores the error structure from the structure of the error support, modeled by morphological graph, and the structure of the error distribution, modeled by exponential distribution. Applying the error structure, Section III proposes the SSEC model and develop an alternately iterative algorithm for its optimization. In Section IV, we verify the proposed method with extensive experiments on popular face databases, and compare it with five state-of-the-art FR techniques. Finally, we give the conclusion and discuss some future work in Section V.

II. Error Structure

In this section, we explore the structure of the error from two aspects: the structure of the error support built on a new graph model, and the structure of the error distribution under robust error metric.

A. Morphological Graph Model for Error Support Structure

1) Morphological Graph Model: We first study the intrinsic error structure from the cognitive style of the visual system of human beings. Why could human beings recognize a facial occlusion accurately and rapidly? Our visual experience indicates that we recognize the occlusion according to its region shape or profile even without knowing what the occlusion is and where it lies. Mathematical morphology shows that the shape can be described by boundaries, skeleton, or the convex hull [27, Chap. 9]. In this work, we try to describe occlusion shape by its boundary, which means the interrelationships between image pixels should be well modeled. While the error support s_i can be used to indicate if an image pixel y_i is occluded, s_i cannot be used to infer if the image pixels around y_i are occluded. Inspired by the work of [25], Zhou *et al.* [9] suggest to model the error support s using a graph $G = (V, E)$. Here, $V = \{1, \ldots, m\}$ denotes the vertex set of m pixels of $y \in \mathbb{R}^m$ and each vertex $i \in V$ is labeled by s_i; $E = E(r) = \left\{(i, j) \,|i, j \in V, \left\|c_i - c_j\right\|_2 \leq r\right\}$ denotes the edges connecting neighboring pixels, where r is the maximum edge length and $c_i = [c_{i1}, c_{i2}]^T$ is the coordinate vector of the vertex i[4]. Thus, the occlusion probabilities of the image pixels around y_i can be inferred from s_i according to the edges connected with vertex i. Clearly, the graph G defined above cannot describe the occlusion shape. Since the occlusion on an image corresponds to the subgraph with vertices labeled by $s_i = 1$, we then consider how to represent the shape of a subgraph.

Let L_G denote the label set of the graph G. Then, G can be divided into several subgraphs by L_G: $G = \{G_k | G_k = (V_k, E_k), k \in L_G\}$, where $V_k = \{i | i \in V, s_i = k\}$ and $E_k = \{(i, j) \,|i, j \in V_k, (i, j) \in E\}$. Since the shape can be described by boundaries, we incorporate the boundary set B of all subgraphs into the classical graph model to form a new graph model $\mathcal{G} = (V, E, B)$, dubbed the *morphological graph*. Before defining the boundary B, we first introduce two special vertex sets to describe the relationships between vertices: the outside vertex set and the related vertex set. The *outside vertex set* v_i^o of the vertex i consists of the vertices which are connected but have different labels with the vertex i: $v_i^o = \left\{j \,|\, (i, j) \in E, s_i \neq s_j\right\}$. The *related vertex set* v_{ij}^o of the vertices i and j consists of the vertex pairs which are connected by the edge but belong to v_i^o and v_j^o, respectively: $v_{ij}^o = \left\{k, l \,|\, (k, l) \in E, k \in v_i^o, l \in v_j^o\right\}$. We then define the boundary set B of the morphological graph $\mathcal{G}$ as following:

Definition 1: The boundary set B of the morphological graph $\mathcal{G} = (V, E, B)$ is the set of the boundaries of all subgraphs of $G = (V, E)$: $B = \{B_k | k \in L_G\}$, where B_k is the boundary of G_k. B_k is also a graph $B_k = \left(\bar{V}_k, \bar{E}_k\right)$, where $\bar{V}_k = \left\{i | i \in V_k, v_i^o \neq \emptyset\right\}$ and $\bar{E}_k = \left\{(i, j) \,|i, j \in \bar{V}_k, v_{ij}^o \neq \emptyset\right\}$.

Section B of the supplementary appendix explores in detail how to detect the subgraph boundaries for a given graph G and how the maximum edge length r of G affects its discriminability. Since subgraphs with $r = 1$ seems to have better discriminability than $r = 2$, we choose $r = 1$ in the following work. In the next subsection, we will show how to use the morphological graph to describe the priori information of the error support s.

2) Priori Probability for Error Support: We now consider how to build the prior probability $p(s)$ of the error support s on the morphological graph $\mathcal{G} = (V, E, B)$. In [9], $p(s)$ is built on the graph $G = (V, E)$ using the classical *Ising model*:

$$p(s) \propto \exp\left(\lambda_E \sum_{(i,j)\in E} s_i s_j + \lambda_V \sum_{i \in V} s_i\right), \tag{2}$$

where the smooth cost $\lambda_E \sum_{(i,j)\in E} s_i s_j$ ($\lambda_E \geq 0$) describes the continuity of the occlusion, and the data cost $\lambda_V \sum_{i\in V} s_i$ ($\lambda_V \geq 0$) gives the priori assumptions about the locations of the erroneous pixels. Comparing with G, $\mathcal{G}$ is equipped with an additional graph B. Since the vertices in B are sensitive to the change of the subgraphs (occlusions), we try to *weaken* the influence of these vertices in the Ising model (2). Then, we have the priori probability $p(s)$ of the error support s on $\mathcal{G}$ (see Section C of the supplementary appendix for a proof):

$$p(s) \propto \exp\left(\lambda_E \sum_{(i,j)\in E} s_i s_j + \lambda_V \sum_{i \in V} s_i - \lambda_B \sum_{i \in \bar{V}_1} s_i\right). \tag{3}$$

Note that the boundary cost $\lambda_B \sum_{i\in\bar{V}_1} s_i = \lambda_B \left|\bar{V}_1\right|$ actually reflects the regularity of the occlusion boundary. Actually, for $\forall G \in \mathbb{G}^1$ and $\forall (i, j) \in \bar{E}_k$, we have $\left\|c_i - c_j\right\|_2 \leq \sqrt{2}$, and $\bar{R}_k = \sum_{(i,j)\in\bar{E}_k} \left\|c_i - c_j\right\|_2 \approx \left|\bar{E}_k\right| \approx \left|\bar{V}_k\right|$.

Although we weaken the influence of the vertices on the graph boundaries in (3), it does not mean that they are not important. On the contrary, they are very important in the quality assessment model of Section III-B3.

B. Exponential Probabilitic Model for Error Distribution Structure

When there are occlusions, Yang *et al.* [11], [19] indicate that the error $\hat{e}$ may be far from any specific distribution. Nevertheless, with the assumption that the unoccluded region of the test image y can be approximated sufficiently by the corresponding regions of the training samples, if the error support s is known and the error $\hat{e}$ is coded and measured in a specific way, we argue that $\hat{e}$ might follow a specific distribution. The details about the error coding and error metric are discussed in Section D of the supplementary appendix. In this work, we pay attention to the local error metric $LEM\left(y_i, \hat{y}_i\right) \triangleq h(0) - h\left(y_i - \hat{y}_i\right)$, where $h(x) = \exp\left(-|x|/\sigma\right)$, and the dilation invariance metric $DIM\left(y_i, \hat{y}_i\right) \triangleq \left|\log\left(y_i/\hat{y}_i\right)\right|$. By combining LEM and DIM together, we form a new error

[4]We use the same spatial coordinate convention as [27, Sec. 2.4.2].

metric:

$$LD\left(y_i, \hat{y}_i\right) \triangleq LEM\left(DIM\left(y_i, \hat{y}_i\right), 0\right) = LEM\left(\log y_i, \log \hat{y}_i\right).$$

Although the global error $\hat{e}$ measured by LD seems to have no definite distribution (refer to Fig. 14(f) in Section D of the supplementary appendix), the locality and the dilation invariance of LD induce that the unoccluded-part error $\dot{e} = \dot{s} \odot \hat{e}$, where $\dot{s} = (1-s)/2$, and the occluded-part error $\ddot{e} = \ddot{s} \odot \hat{e}$, where $\ddot{s} = 1 - \dot{s}$, approximately follow independent exponential distributions, respectively. For the known error support s, we therefore suggest the following conditional probability density of the error $\hat{e}$:

$$p\left(\hat{e}|s\right) = p\left(\dot{e}, \ddot{e}|s\right) = p\left(\dot{e}|s\right) p\left(\ddot{e}|s\right),$$

where $p\left(\dot{e}|s\right)$ and $p\left(\ddot{e}|s\right)$ are formulated as

$$p\left(\dot{e}|s\right) = \prod_{i \in V} \left(\dot{\lambda} \exp\left(-\dot{\lambda}\dot{e}_i\right)\right)^{\dot{s}_i}, \tag{4}$$

$$p\left(\ddot{e}|s\right) = \prod_{i \in V} \left(\ddot{\lambda} \exp\left(-\ddot{\lambda}\left(\ddot{s}_i - \ddot{e}_i\right)\right)\right)^{\ddot{s}_i}. \tag{5}$$

The rate parameters $\dot{\lambda}$ and $\ddot{\lambda}$ can be set by their maximum likelihood estimates $\hat{\dot{\lambda}} = ML\left(\dot{s}, \hat{e}\right) = \left(\sum_{i \in V} \dot{s}_i\right) / \left(\sum_{i \in V} \dot{s}_i \hat{e}_i\right)$ and $\hat{\ddot{\lambda}} = ML\left(\ddot{s}, e\right) = \left(\sum_{i \in V} \ddot{s}_i\right) / \left(\sum_{i \in V} \ddot{s}_i \left(1 - \hat{e}_i\right)\right)$, respectively.

Although the probability models of (4) and (5) are based on the assumption that the error support s is exactly known, experiments show that they can also be applied to the iterative procedure where the error support is estimated more and more precisely.

III. Structured Sparse Error Coding

In this section, we first propose the Structured Sparse Error Coding (SSEC) model for FR with occlusion, and then we present an alternately iterative algorithm to solve the SSEC. In our SSEC model, we assume that the pixel values of the test image y and the training set A are all normalized to the interval $[1, 256]$.

A. Proposed Model

Section II explores the error structure by analyzing the structure of the error support and the structure of the error distribution. Using the structured prior information, we consider the joint probability $p(s, e)$:

$$p\left(s, \hat{e}\right) = p\left(s, \dot{e}, \ddot{e}\right) = p\left(\dot{e}, \ddot{e}|s\right) p(s) = p\left(\dot{e}|s\right) p\left(\ddot{e}|s\right) p(s). \tag{6}$$

Substituting (4), (5), (3) into (6), we have

$$p\left(s, \hat{e}\right) \propto \exp\left(\sum_{i \in V} \left(\ddot{\lambda}\ddot{s}_i - \dot{\lambda}\dot{s}_i\right) \hat{e}_i + \lambda_E \sum_{(i,j) \in E} s_i s_j + \log\left(\dot{\lambda}\right) \sum_{i \in V} \dot{s}_i + \left(\log\left(\ddot{\lambda}\right) - \ddot{\lambda}\right) \sum_{i \in V} \ddot{s}_i + \lambda_V \sum_{i \in V} s_i - \lambda_B \sum_{i \in \bar{V}_1} s_i\right). \tag{7}$$

Since the higher the joint probability $p\left(s, \hat{e}\right)$, the higher the probability of the estimated error support coincides with the real error support, we formulate the following optimization model for occlusion detection:

$$\left(s^*, e^*\right) = \arg\max_{s,e} p(s, e) \quad s.t.\ e = LD\left(y, \hat{y}\right), \hat{y} = Ax, x \geq 0. \tag{8}$$

Here, we use the non-negative sparse coding (NNSC) to calculate the coefficient of y w.r.t. A (please see Section D of the supplementary appendix for more details). Since the error e in (8) is coded with structured sparse representation, we refer the optimization problem (8) to *Structured Sparse Error Coding* (SSEC).

B. Algorithm of SSEC

Since there are two variables e and s in the objective function of the SSEC model (8), (8) can be solved in an alternating maximization way

$$e^{(t)} = \arg\max_{e} p\left(s^{(t-1)}, e\right) \quad s.t.\ e = LD\left(y, \hat{y}\right), \hat{y} = Ax, x \geq 0, \tag{9}$$

$$s^{(t)} = \arg\max_{s} p\left(s, e^{(t)}\right). \tag{10}$$

Section III-B.1 will probe how to recover the error $e^{(t)}$ by iteratively reweighted sparse coding [10], [11]. Section III-B.2 will probe how to recover the error support $s^{(t)}$ by combining the threshold clustering and the graph cuts clustering [28].

1) Recovering Error by Iteratively Reweighted Sparse Coding: Since $e = LD(y, Ax)$, $x \geq 0$, for a fixed $s = s^{(t)}$, (9) can be reformulated as

$$x^* = \arg\max_{x} \sum_{i \in V} \left(\dot{\lambda}\dot{s}_i - \ddot{\lambda}\ddot{s}_i\right) h\left(\log y_i - \log \sum_j a_{ij} x_j\right) \quad s.t.\ x \geq 0, \tag{11}$$

where a_{ij} is the jth entry of the training sample a_i. Now, the optimization problem used to recover the error e is converted to recover the non-negative sparse coding x. Since the objective function in (11) is similar to the objective function of CESR in [10], the half-quadratic technique [29] and expectation maximization (EM) method can also be utilized to solve this optimization problem. The detailed derivations can be found in Section E of the supplementary appendix. We summarize the derived results into the following weighted sparse error coding model:

$$w^{(t)} = \dot{s} \odot h\left(\left|\log y - \log Ax^{(t-1)}\right|\right), \tag{12}$$

$$x^{(t)} = \arg\min_{x} \left\| w^{(t)} \odot (y - Ax) \right\|_1 \quad s.t.\ x \geq 0. \tag{13}$$

Now, for the fixed error support s, using (12) and (13), the weight $w^{(t)}$ and the sparse coefficient $x^{(t)}$ can be updated alternately until convergence. In essence, the whole alternative procedure is the iteratively reweighted sparse coding (IRSC) [11], [30]. For convenience, we write the iterative result $\hat{x}$ as $\hat{x} = \mathbf{IRSC}(A, y, s)$, and then the recovered error $\hat{e} = LD\left(y, A\hat{x}\right)$.

2) Recovering Error Support by Triple Clustering: Since $\dot{s} = (1-s)/2$, $\ddot{s} = (1+s)/2$, for the fixed error $e = e^{(t)}$, (10) can be reformulated as

$$s^* = \arg\max_s S_c(s, \lambda_E) + D_c\left(s, \dot{\lambda}, \ddot{\lambda}, \lambda_V\right) + B_c\left(s, \lambda_B, \bar{V}_1\right), \tag{14}$$

where $S_c(s, \lambda_E) = \lambda_E \sum_{(i,j)\in E} s_i s_j$ is the smooth cost, $D_c(s, \dot{\lambda}, \ddot{\lambda}, \lambda_V) = \sum_{i\in V}(\log(\dot{\lambda}) - \dot{\lambda}\dot{e}_i)\dot{s}_i + \sum_{i\in V}(\log(\ddot{\lambda}) - \ddot{\lambda} + \ddot{\lambda}\ddot{e}_i)\ddot{s}_i + \lambda_V \sum_{i\in V} s_i$ is the data cost, and $B_c(s, \lambda_B, \bar{V}_1) = -\lambda_B \sum_{i\in\bar{V}_1} s_i$ is the boundary cost.

The optimization problem (14) is an integer programming problem used to cluster the image pixels into two classes: occluded ($s_i = 1$) and unoccluded ($s_i = -1$). If we combine the boundary cost into the data cost, and then (14) accords with the classical Ising model and can be solved in linear time using graph cuts [28]. However, before directly solving (14) using graph cuts, we should first estimate the error support s and the corresponding boundary vertex set $\bar{V}_1$. A direct way to estimate s is using threshold clustering model

$$s_i = \begin{cases} 1, & |e_i| > \tau \\ -1, & |e_i| \le \tau. \end{cases} \tag{15}$$

Recall that $\bar{V}_1$ can be calculated by the known s (see Section II-A.1). However, the error support estimated by (15) is unsuitable to be used to calculate $\bar{V}_1$, since the spatial continuity is not considered. Hence, we further estimate s by (14) with $\lambda_B = 0$. Based on this reestimated s, we calculate $\bar{V}_1$. Finally, using the reestimated s, $\bar{V}_1$ and the input λ_B, we, once again, estimate s. The procedure used to estimate s or solve (14) can be regarded as clustering the image pixels. Hence, to solve (14), we need triple clustering $s = \mathbf{TC}(e, \tau, \lambda_E, \lambda_V, \lambda_B)$. Note that the threshold τ is actually an initial decision boundary to separate the pixels into occluded and unoccluded ones. A small τ might result in many false positives (unoccluded pixels are falsely labeled as occluded) while a large τ might result in many false negatives. We therefore adopt an *increasing-after-decreasing* strategy to adjust τ during the iterative procedure, i.e., we let $\tau = [\tau^{(0)}, \tau^{(0)} - \delta, \ldots, \tau^{(0)} - (T-1)\delta, \tau^{(0)} - (T-2)\delta, \ldots, \tau^{(0)} - \delta, \tau^{(0)}]$, where $\delta = -((k-1)/(T-1))\tau^{(0)}$. Please see Section F of the supplementary appendix for more details. Fig. 2 in Section IV-A.1 shows the effect of the threshold adjusting strategy.

3) Quality Assessment Model: With iterations, our algorithm produces a sequence of recovered images $\{\hat{y}^{(t)} = Ax^{(t)}, t = 1, 2, \ldots\}$, and a sequence of estimated error supports $\{\hat{s}^{(t)}, t = 1, 2, \ldots\}$. A direct way to assess the quality of these recovered images and error supports is to calculate the total energy $E(s^{(t)}, \dot{e}^{(t)}, \ddot{e}^{(t)}) = \log(p(s^{(t)}, \dot{e}^{(t)}, \ddot{e}^{(t)}))$ and choose the maximum one. However, the threshold sequence τ introduced into the model might cause that a maximum total energy corresponds to the worst estimates. In order to reduce the effect of the threshold τ, we just try to maximize part terms (the unoccluded term and the boundary term) of the total energy $E(s^{(t)}, \dot{e}^{(t)}, \ddot{e}^{(t)})$, i.e., we try to find

$$t^* = \arg\max_t \sum_{i\in V}\left(\log\left(\dot{\lambda}\right)\dot{s}_i^{(t)} - \dot{\lambda}\dot{e}_i^{(t)}\right) - \lambda_B \sum_{i\in\bar{V}_1} s_i^{(t)}. \tag{16}$$

Substituting $\dot{\lambda}$'s maximum likelihood estimate $\hat{\dot{\lambda}} = ML(\dot{s}, e)$ into (16), we have

$$t^* = \arg\min_t c_E^{(t)} + \lambda_B c_B^{(t)}, \tag{17}$$

where $c_E^{(t)}$ is the energy of the reconstructed errors in the unoccluded area:

$$c_E^{(t)} = \left(1 + \log\left(\frac{\sum_{i\in V}\dot{e}_i^{(t)}}{\sum_{i\in V}\dot{s}_i^{(t)}}\right)\right)\sum_{i\in V}\dot{s}_i^{(t)}, \tag{18}$$

and $c_B^{(t)}$ is the boundary term

$$c_B^{(t)} = \sum_{i\in\bar{V}_1} s_i^{(t)}. \tag{19}$$

Clearly, it's very important to estimate the boundary regularization parameter λ_B. For convenience, we first normalize the energy term series $\{c_E^{(t)}\}_t$ and the boundary term series $\{c_B^{(t)}\}_t$ to be in the range $[0, 1]$, respectively. We can estimate λ_B by cross-validation or by experience. However, these methods are too complex or too subtle to be used in practice. Here, we suggest an automatic way to estimate λ_B. Notice that $c_E^{(t)}$ and $c_B^{(t)}$ interact each other: if the detected occluded area is big, $c_B^{(t)}$ might be small because the scattered occluded patches will be integrated to one whole block; however, a big occluded area will lead to a small unoccluded area, $\sum_{i\in V}\dot{s}_i^{(t)}$, which might lead to a big $c_E^{(t)}$. Intuitively, the numerical stability of the energy term series $\{c_E^{(t)}\}_t$ and the boundary term series $\{c_B^{(t)}\}_t$ measures their confidence levels (although, in our view, the numerical stability is not suitable for determining convergence), i.e., which series is better for quality assessment. Since the differences of a series reflect its numerical stability, we suggest to estimate λ_B by

$$\hat{\lambda}_B = \frac{\sum_t \left|c_E^{(t+1)} - c_E^{(t)}\right|}{\sum_t \left|c_B^{(t+1)} - c_B^{(t)}\right|}. \tag{20}$$

4) Final Algorithm: Incorporating the procedures of recovering the test image, detecting the occlusion, and assessing the quality of the recovered images and error supports, Algorithm 1 summarizes the complete procedure used to solve the SSEC model (8).

IV. Simulations and Experiments

To evaluate the proposed SSEC algorithm, we compare it with the state-of-the-art methods for FR with occlusions. Except for the error coding-based methods we reviewed in the introduction: CESR[5] [10], RSC[6] [11] and SEC_MRF [9], another two skin color-based methods (i.e., the occlusion is detected by skin color not by the reconstruction error): Partial

[5]The Matlab source code: http://www.openpr.org.cn/index.php/All/69-CESR/View-details.html.

[6]The Matlab source code: http://www4.comp.polyu.edu.hk/ cslzhang/code.htm.

Algorithm 1 Structured Sparse Error Coding (SSEC)

Input: data matrix A, test sample y.

Output: estimated error $\hat{e}$, estimated error support $\hat{s}$, identity (y)

1) Initialize error $[e^{(0)}, x^{(0)}] = \mathbf{IRSC}\,(A, y, -\mathbf{1})$;
2) Initialize error support $s^{(0)} = \mathcal{K}\left(e^{(0)}\right)$;
3) Initialize the threshold sequence τ according to the strategy suggested in Section F of the supplementary appendix;
4) **For** t = 1 **to** $2T-1$
5) Error recovering: $\left[e^{(t)}, x^{(t)}\right] = \mathbf{IRSC}\left(A, y, s^{(t-1)}, 1\right)$;
6) Error support recovering: $s^{(t)} = \mathbf{TC}\left(e^{(t)}, \tau_t, \lambda_E, \lambda_V, \lambda_B\right)$;
7) Calculate $c_E^{(t)}$ and $c_B^{(t)}$ according to (18) and (19), respectively;
8) **End For**
9) Normalize $c_E^{(t)}$ and $c_B^{(t)}$ to be in the range $[0, 1]$;
10) Assess the quality: $t^* = \arg\min_t c_E^{(t)} + \lambda_B c_B^{(t)}$, where $\lambda_B = \frac{\sum_t \left| c_E^{(t+1)} - c_E^{(t)} \right|}{\sum_t \left| c_B^{(t+1)} - c_B^{(t)} \right|}$;
11) Set $\hat{s} = \dot{s}^{(t^*)}$, $\hat{x} = x^{(t^*)}$, $\hat{e} = y - A\hat{x}$;
12) For $k = 1, \ldots, K$, compute the residuals $r_k = \left\| \hat{s} \odot LD\left(y - A\delta_k\left(\hat{x}\right)\right) \right\|_1$, where $\delta_k\left(\hat{x}\right)$ is a new vector whose only nonzero entries are the entries in $\hat{x}$ that are associated with class k;
13) identity $(y) = \arg\min_k r_k$.

Within-Class Match (PWCM) [5] and Partial Support Vector Machines (PSVM) [6][7] are also contained in the following experiments. For PWCM, we choose the ℓ_1-norm to calculate its reconstruction error for classification and denote it as PWCM_1. For PSVM, according to the provided Matlab code, we just use the linearly separable PSVM (see [6, Eq. (7)]). The parameters of the other methods are set according to the strategy suggested in their papers. The parameters of our SSEC algorithm are set as: $\lambda_E = 2$, $\lambda_V = 0$, $\kappa = 0.3$, and $T = 5$. The boundary regularization parameter λ_B is set to 0 for solving (14). In order to explore the effect of the boundary term in (17), we perform quality assessment with different values of $\lambda_B \in \{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 50, \hat{\lambda}_B\}$, where $\hat{\lambda}_B$ is automatically calculated by (20). We use SSEC_Optimal and SSEC_Auto to denote SSEC with λ_B optimally chosen from its parameter set and SSEC with λ_B calculated by (20), respectively. The K-means operator $\mathcal{K}(\cdot)$, used to initialize $\tau^{(0)}$ and $s^{(0)}$, adopts the sum-of-absolute-error criterion as its distance measure.

Experiments are performed on two publicly available databases, namely, the Extended Yale B [15], [31] database and the AR [32] database. All algorithms are implemented in Matlab on a Dual-Core Intel Core i5-2300 CPU 2.80GHz computer with 8.00 GB Memory (RAM) and Windows Server 2008 64-bit Operating System.

Based on the experimental results on the two publicly available databases, we try to draw the following conclusions about SSEC:

1) The boundary feature is also an important feature especially when the occlusion level is high, or the feature dimension is low. Using the boundary regularized quality assessment model (17), SSEC achieves much more stable recognition performance for various levels of occlusion and various feature dimensions than the other compared methods.
2) With feature dimension increasing, the boundary regularization parameter λ_B calculated by (20) approximates its optimal estimate; if the feature dimension is high enough, λ_B calculated by (20) approximates its optimal estimation for almost all occlusion levels.

A. Synthetic Occlusion With the Extended Yale B Database

In this section, we use the Extended Yale B database [15], [31] to investigate the breakdown point and the effect of the boundary term (19) of SSEC under fixed feature dimension for various synthetic occlusions with varying levels and varying boundaries. The Extended Yale B database consists of 2,414 frontal-face images of 38 individuals under 64 laboratory-controlled illumination conditions [15], [31]. According to the illumination conditions, the images have been divided into 5 subsets [15]. Among the 1,174 face images in Subset I (mild illuminations), Subset II (moderate illuminations), and Subset III (heavy illuminations), we randomly choose 4 illuminations from Subset I , 2 from Subset II, and 2 from Subset III for testing, and the rest for training. The total numbers of images in the training and test sets are 870 and 304, respectively. The images are cropped to 96×84 pixels. In the following experiments, synthetic occlusions with various boundaries and various occlusion levels (0∼90%) are imposed on the test samples. Note that this experimental setting is identical to the one in [9, Sec. 4] and satisfies the preconditions we assume in the introduction (please see Section G of the supplementary appendix for details).

1) Recognition With Straight-Line-Boundary Occlusion: Training and test sets are chosen according to the scheme described at the beginning of this section. To compare the breakdown point of SSEC with the other competitive algorithms, we simulate various levels of contiguous occlusions from 0% to 90% by replacing a random located block of each test image with a mandrill image. Fig. 1(a) shows the recognition performances of different methods. If we define the breakdown point of an algorithm as the highest occlusion level that the algorithm can tolerate when the recognition rate is required to be not less than 90%, then according to Fig. 1(a), the breakdown points of CESR, RSC, SEC_MRF, PSVM, PWCM_1, and SSEC are 60%, 40%, 80%, 20%, 80%, and >90%, respectively. The Precision-Recall (PR) curves for 80% and 90% occlusions are also plotted in Fig. 1(b) and (c), respectively. PR is an important index for evaluating FR methods against outliers. A good algorithm should achieve a high precision ratio at a high recall ratio.

As seen from the recognition rate curves and the PR curves, RSC has more robust recognition performance than CESR when the occlusion level is high, although RSC has lower breakdown point (40%) than CESR (60%). Different with

[7]The Matlab source code: http://cbcsl.ece.ohio-state.edu/downloads.html.

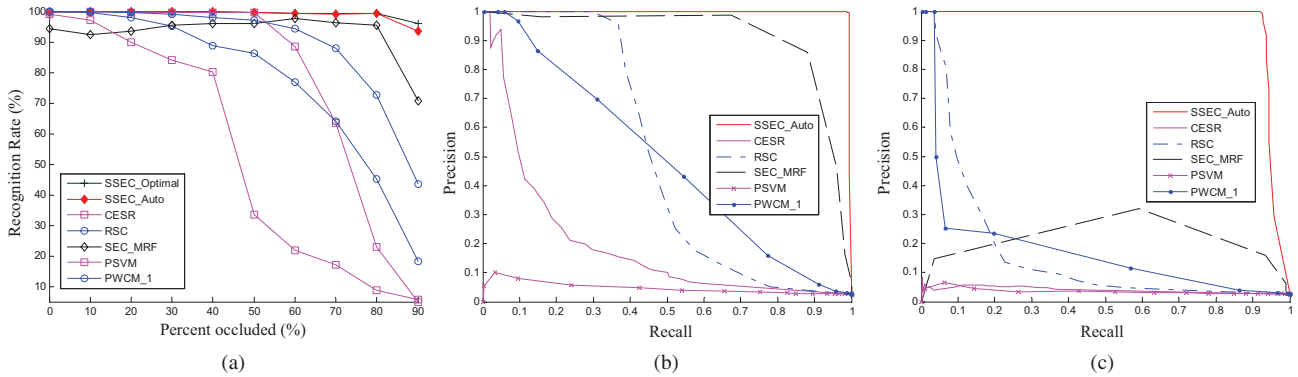


Fig. 1. Recognition with mandrill occlusion on the extended yale B database. (a) Recognition rates for various algorithms against 0–90% mandrill occlusions. (b) PR curves for 80% occlusion. (c) PR curves for 90% occlusion.

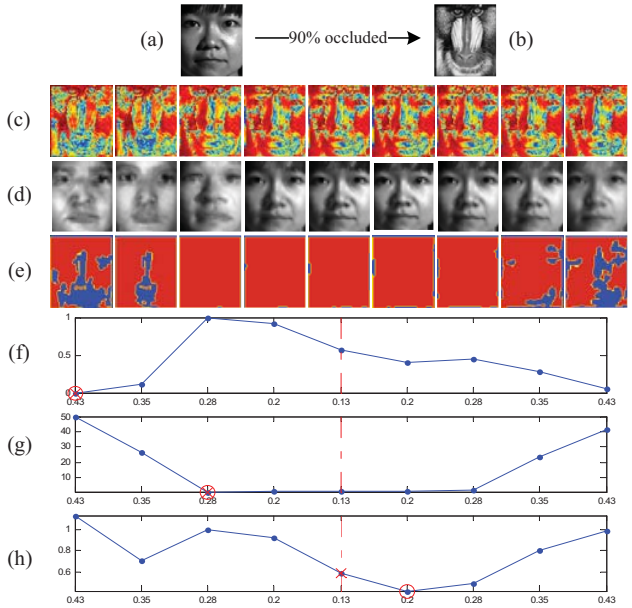


Fig. 2. Quality assessment of SSEC for a test image with 90% occlusion. (a) Ground truth face image from subject 5 of the extended yale B database. (b) 90% mandrill occluded image. (c)–(e) Series of the estimated errors, the recovered images, and the estimated error supports (red square box: occluded; blue square box: unoccluded; yellow square box: boundary) produced during the SSEC iterations. (f)–(h) Quality assessment according to (17) with different regularization values: $\lambda_B = 0$, $\lambda_B = 50$, and λ_B calculated according to (20). When $\tau^{(0)}$ decreases to $\kappa\tau^{(0)}$, the minimum point is marked by red cross symbol. For all involved threshold τ's, the global minimum point is marked by red ring.

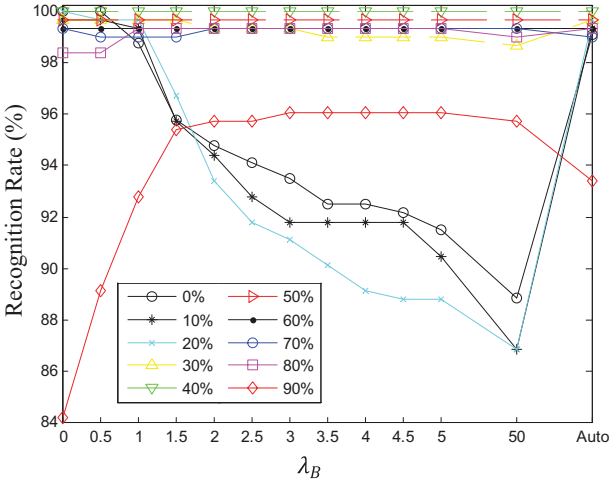


Fig. 3. Recognition results of SSEC with various values of λ_B under various levels of mandrill occlusions (with straight line boundary).

Fig. 4. Occlusions with various boundaries. (a) *Mandrill*. (b) *Camera*. (c) *Dog*. (d) *Apple*. (e) *Flower*.

CESR and RSC, the recognition rates of SEC_MRF[8] are much more stable for various occlusion levels, except the 90% occlusion. For the 90% occlusion, although SEC_MRF has a much higher recognition rate than RSC, it has much lower precision ratio than RSC when the recall ratio is less than 0.2. This means that the recognition performance of SEC_MRF becomes unstable for too high occlusion levels. For PWCM_1 and PSVM, suprisingly, although they both use the same occlusion detector (skin color detector), their recognition rates are terribly different. The main reason might be that: 1) the detected occlusion regions can be conveniently discarded by PWCM_1, while they have to be recovered by PSVM; 2) the strict linearly separable constraint of PSVM might result in overfitting for the images with so high dimension ($96 \times 84 = 8064$).

Compared to the other 5 methods, both SSEC_Auto and SSEC_Optimal have high breakdown points and perform well almost for each occlusion level. Specially, for the extreme circumstance (the 90% occlusion), although the occluded images (such as the image in Fig. 2(b)) are barely recognizable

[8]The recognition rates of SEC_MRF reported here are taken from our own experimental results. For the 90% occlusion, the recognition rate of SEC_MRF reported in [9] is 40.3%, which is much lower than the one, 70.7%, reported here. The main reason might be that for our own Matlab implementation of SEC_MRF, the threshold $\tau^{(0)} = 0.17$ we choose is more suitable for high level occlusion than for low level occlusion. For the same reason, the recognition rates of SEC_MRF reported here are roughly increasing with the occlusion level when the occluded percent is less than 60% (for the occlusion levels from 10–60%, the recognition rates reported in [9] are all 100%).

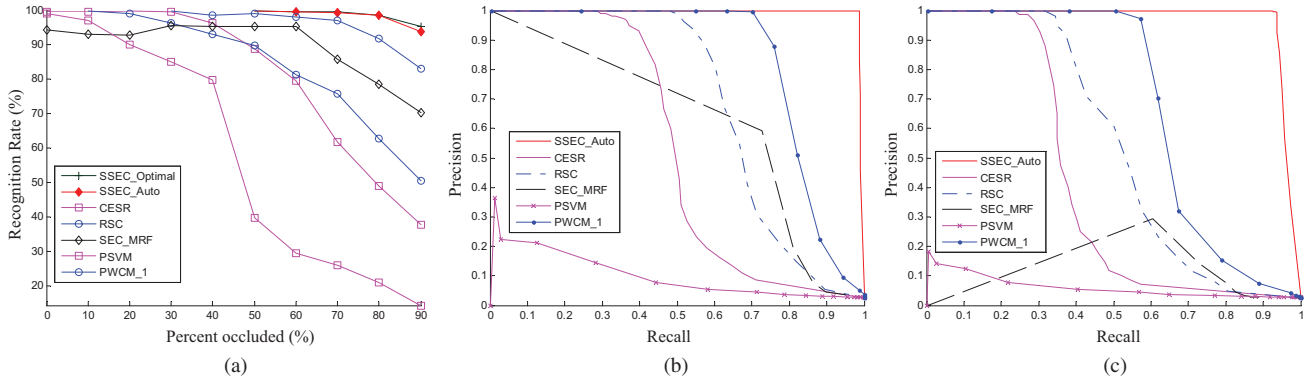


Fig. 5. Recognition with various boundary occlusions on the extended yale B database. (a) Recognition rates for various algorithms against 0–90% various occlusions. (b) PR curves for 80% occlusion. (c) PR curves for 90% occlusion.

as face images, the recognition rate of SSEC is still larger than 90%. The astonishing high recognition rate indicates that SSEC could detect occlusions accurately. Fig. 2 shows how SSEC works in this extreme circumstance. The curves plotted in Fig. 2(f)–(h) illustrate how SSEC chooses the well estimated variables according to (17) with various boundary regularization values. Compared to $\lambda_B = 0$ (Fig. 2(f)) and $\lambda_B = 50$ (Fig. 2(g)), λ_B calculated according to (20) leads to better recovered images and error supports.

In order to explore the effect of the quality assessment model (17), we further show the recognition results of SSEC for various values of λ_B under various occlusion levels in Fig. 3. With λ_B increasing, the recognition rates drop significantly when the occlusion level is less than 30%, drop gently when the occlusion level is equal to 30%, and almost do not change when the occlusion level increases from 40–70%. The increasing values of λ_B begin to enhance the recognition rates when the occlusion level achieves 80% and, especially, 90%. It means that a good choice of λ_B also depends on the occlusion level, which, unfortunately, is usually unknown a priori. Fortunately, as seen from Fig. 3, for all occlusion levels, except 90%, SSEC_Auto has similar recognition rates with SSEC_Optimal.

2) Recognition With Various-Boundary Occlusion: Intuitively, the constraint of the minimum boundary perimeter in the SSEC model (8) might result in SSEC having poor performances in dealing with occlusions with curving boundaries. In this section, we simulate occlusions with various boundaries, as shown in Fig. 4. Training and test sets are chosen according to the same scheme as Section IV-A.1.

The recognition results of the compared methods are reported in Fig. 5. Not surprisingly, the recognition rates of CESR, RSC, and PWCM_1 are greatly improved in this experiment for the high occlusion levels (exceeds 70%), since the pixels of the occlusion images used in this setting are more uncorrelated with the pixels of face images than the mandrill image used in Section IV-A.1. The recognition rates of SEC_MRF keep at the same level for 90% occlusion in these two experiments while its breakdown point decreases to 60% occlusion in this experiment, which means, for SEC_MRF, the occlusion shapes affect its recognition rates more than the occlusion pixel values. Surprisingly, the occlusion shapes seem

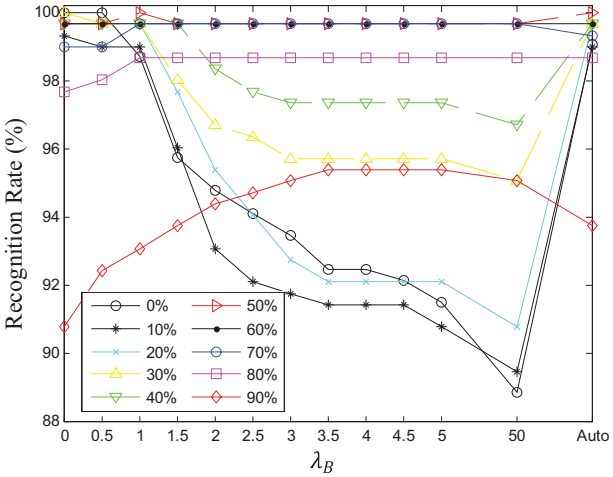


Fig. 6. Recognition rates of SSEC with various values of λ_B under various levels of occlusions with various boundaries.

not to affect the recognition performances of SSEC. Part of the reasons might be that compared with the irregular boundary of the wrongly estimated occlusion, the boundary of the natural occlusion image looks much more regular. Some examples of the detected occlusions of SSEC are shown in Section H1 of the supplementary appendix.

Fig. 6 explores the effect of the quality assessment model (17) of SSEC for various values of λ_B under various occlusion levels. Although SSEC is very sensitive to the occlusion boundary when the occlusion percent is less than 50%, SSEC_Auto (λ_B calculated by (20)) approximates the optimal performances under all occlusion levels, except 90% occlusion.

B. Real-World Occlusion With the AR Database

We test the ability of SSEC in dealing with real disguises using the AR face database [32], which is commonly used for evaluating robust FR algorithms. The AR database consists of over 4,000 facial images from 126 subjects (70 men and 56 women). For each subject, 26 facial images are taken in two separate sessions. These images suffer different facial variations including various facial expressions (neutral, smile, anger, and scream), illumination variations (left light on, right

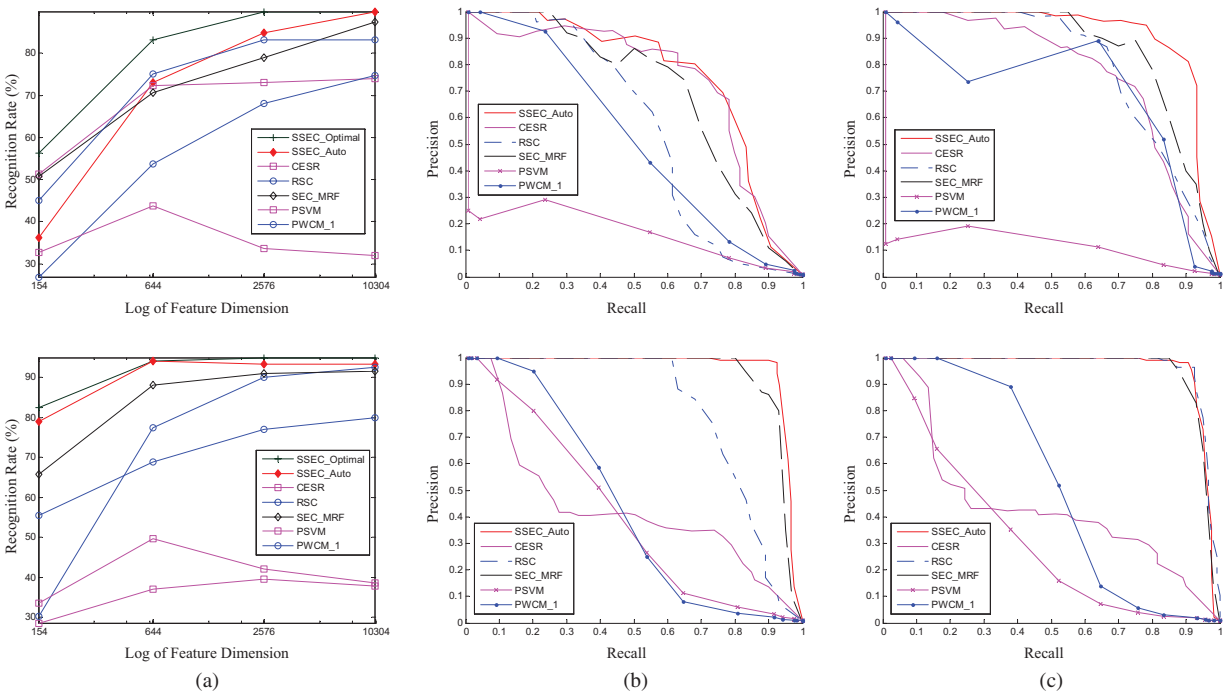


Fig. 7. Recognition rates and PR curves of various algorithms under various feature spaces against sunglasses and scarf occlusions in the AR database. Top row: sunglasses occlusion. Bottom row: scarves occlusion. (a) Recognition rates. (b) PR on dimension 644. (c) PR on dimension 1,0304.

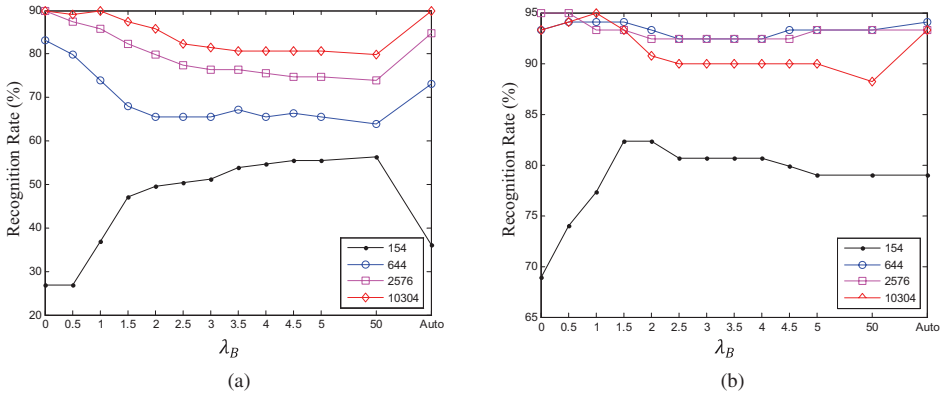


Fig. 8. Recognition results of SSEC with various values of λ_B and various feature dimension. (a) Sunglasses occlusion. (b) Scarf occlusion.

light on and all side lights on), and occlusions (sunglasses and scarves). In this section, we selected a subset of the database that consists of 119 subjects (65 males and 54 females). The grayscale images were resized to resolution 112×92.

1) Recognition Without Occlusions in the Training Set: For training, we choose 8 unoccluded frontal view images with varying facial expressions for each of 119 subjects from the first and second session. For testing, we consider two separate test sets of the 119 subjects. The first test set contains 119 images of the subjects wearing sunglasses, which occlude roughly 20% of the image. The second test set contains 119 images of the subjects wearing a scarf, which occludes roughly 40% of the image.

Fig. 7 compares the FR performance of different methods using 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. Generally speaking, recognition rates increase with feature dimensions increasing. However, the overfitting problem induces that the linearly separable PSVM does not achieve its optimal recognition rate

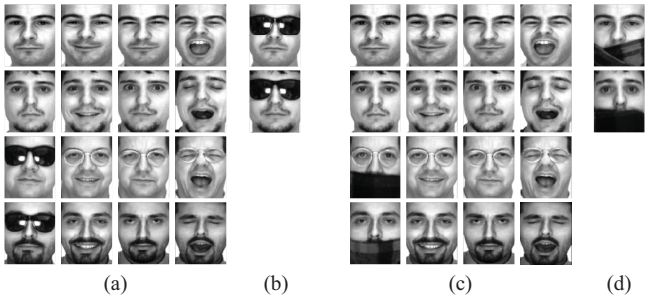


Fig. 9. Examples of the training and test sets used in Section IV-B.2 for $|A| = |B| = 2$. (a) Training set partially occluded by sunglasses. (b) Test set with sunglasses occlusion. (c) Training set partially occluded by scarf. (d) Test set with scarf occlusion.

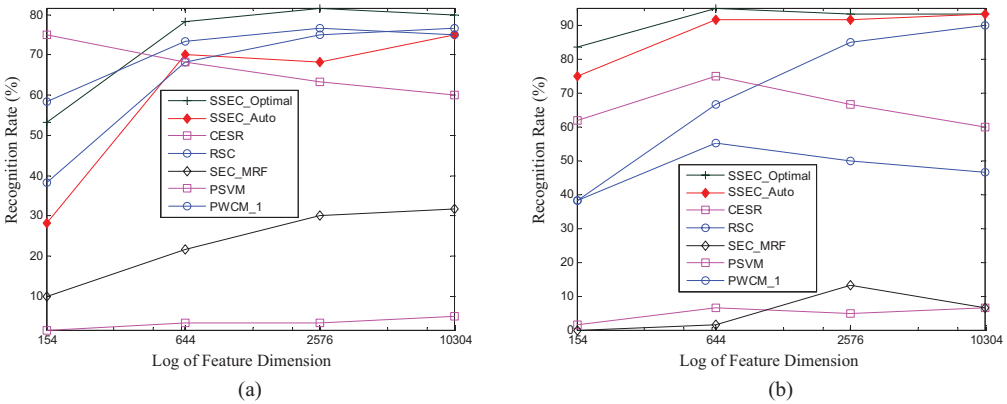


Fig. 10. Recognition rates of various algorithms under various feature spaces against sunglasses and scarves occlusions. (a) Sunglasses occlusion. (b) Scarves occlusion.

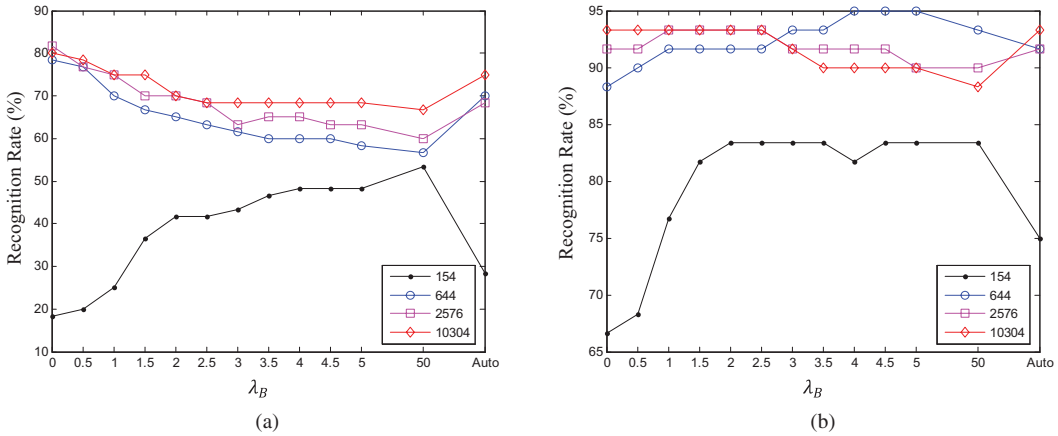


Fig. 11. Recognition results of SSEC with various values of λ_B and various feature dimension. (a) Sunglasses occlusion. (b) Scarf occlusion.

at the highest dimension 1,0304. The Locally Salient Features (LSF) existing in both the training and test tests (e.g., the moustache in the training set and the scarf in the test set), lead CESR to have much lower recognition rates against scarf occlusion than against sunglasses occlusion. Compared to CESR, however, RSC, SEC_MRF, PSVM and SSEC all work better for the scarf occlusion than for sunglasses occlusion. In spite of this, it does not mean that all of these methods are robust to the LSFs, and the problem of the LSFs will be explored deeply in the next subsection. Here, we just explain the working principle of our SSEC method: for the scarf occlusion, due to its remarkable boundary, both SSEC_Auto and SSEC_Optimal do better than the other methods, especially for the low feature dimension, as illustrated in the bottom row of Fig. 7; for sunglasses occlusion, since its boundary is not as remarkable as the scarf occlusion, SSEC (SSEC_Auto) does not always do better than the other methods, but its recognition performance increases dramatically with the feature dimension increasing (because the sunglasses occlusions have more and more remarkable boundaries with the feature dimension increasing), as illustrated in the top row of Fig. 7. Some examples of the detected occlusions of SSEC are shown in Section H2 of the supplementary appendix.

The recognition performance of SSEC with various feature dimensions shows that: 1) the boundary feature is also an important feature especially when the feature dimension is very low; 2) with feature dimension increasing, the boundary regularization parameter λ_B estimated by equation (20) becomes better and better. These viewpoints are further illustrated in Fig. 8.

2) Recognition With Occlusions in Part of the Training Set: To test the robustness of SSEC against the LSFs, we add occluded samples to *part* of the training set, and simultaneously, reduce the number of the training samples. Specifically, we choose the training and test sets as follows. We first randomly divide all or part of the 119 subjects into subset A and B, and $A \cap B = \emptyset$. Then, we choose the training set from the subjects in A and B, and choose the test set only from the subjects in A. For testing, as well as in Section IV-B.1, we still consider two separate test sets: one only contains images disguised by sunglasses, and the other only contains images disguised by scarves. For training, for each subject in A and B, we select 4 images of unoccluded frontal views with varying facial expressions (neutral, smile, anger, and scream) from the first session; then, for each subject in B, we replace the neutral-expression sample with the sunglasses/scarf-disguised image. For $|A| = |B| = 2$, Fig. 9 shows how the training and test sets are divided. Clearly, the sunglasses/scarf disguises existing in both the training and test sets form the LSFs.

By setting $|A| = 60$ and $|B| = 59$, we choose the training and test sets according to the above subset dividing scheme. Fig. 10 shows the recognition rates of the compared

algorithms under various feature spaces against sunglasses and scarf disguises. Although the maximum correntropy criterion used in CESR is insensitive to outliers, the outliers, such as the sunglasses/scarf disguises, coexisting in both the training and test sets will not be recognized as outliers by the CIM metric, which leads CESR to have low recognition rates. The strategy used by SEC_MRF to identify the test image (i.e., comparing the test image with the training ones of each subject one by one) causes SEC_MRF to be insensitive to the LSFs incurred by occlusion, especially when the occluded regions are large. In this case, RSC is the nearest competitor of SSEC. Due to using the occlusion information in the training set, both PWCM_1 and PSVM do better than SSEC for the low occlusion level and low feature dimension. However, with the occlusion level and the feature dimension increasing, the performances of PWCM_1 and PSVM tend to decline.

The recognition rates of SSEC_Optimal greatly outperform the other compared algorithms, which shows that the proposed structured error model does be insensitive to the LSFs. However, the LSFs coexisting in both the training and test sets disturb the recognition performance of SSEC_Auto for the sunglasses disguise, and make its recognition rates drop when the feature dimension increases from 644 to 2576. This means that the boundary regularization parameter λ_B is not well estimated by (20) in this situation. The curves of the recognition rates versus various λ_B plotted in Fig. 11 show the same regularity as the ones plotted in Fig. 8.

V. Conclusion

In order to efficiently perform FR with occlusion, we explore the intrinsic structure of the error incurred by occlusion from two aspects: the morphological feature and the probabilistic distribution. Based on the morphological graph model and the probabilistic model of the error, we propose the Structured Sparse Error Coding (SSEC) model for FR with occlusion. Experimental results on both synthetic and real-world occlusions show that SSEC has much more stable recognition performance in dealing with occlusions as compared to the state-of-the-art methods. For the quality assessment model of SSEC, the boundary regularization parameter λ_B plays an important role especially for the extreme situation (e.g., the high-level occlusion or the low feature dimension). However, λ_B calculated by equation (20) does not always approximate the optimal one. How to estimate λ_B accurately and automatically will be our future work. In addition, we also expect the SSEC model could be extended to deal with other more complex real-world occlusions, such as different disguises with different colors, and different occlusions disturbed by mixed variations of poses/expressions/illuminations, in future.

Acknowledgment

The authors would like to thank W.-S. Zheng, R. He, and Z.-H. Zhou, who provided the details of their algorithms. They also thank all reviewers and the associate editor for their thorough and helpful comments.

References

[1] D. Hoiem, A. N. Stein, A. A. Efros, and M. Hebert, "Recovering occlusion boundaries from a single image," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[2] A. Humayun, O. M. Aodha, and G. J. Brostow, "Learning to find occlusion regions," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Aug. 2011, pp. 2161–2168.

[3] A. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

[4] S. Fidler, D. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 337–350, Jan. 2006.

[5] H. Jia and A. M. Martinez, "Face recognition with occlusions in the training and testing sets," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Sep. 2008, pp. 1–6.

[6] H. Jia and A. M. Martinez, "Support vector Mach.s in face recognition with occlusions," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 136–141.

[7] X. Y. Tan, S. C. Chen, Z. Z. Zhou, and J. Liu, "Face recognition under occlusions and variant expressions with partial similarity," *IEEE Trans. Inf. Forensics Security*, vol. 4, no. 2, pp. 217–230, Jun. 2009.

[8] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[9] Z. H. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using markov random fields," in *Proc. IEEE Int. Conf. Comput. Vis.*, Sep.–Oct. 2009, pp. 1050–1057.

[10] R. He, W. Zheng, and B. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[11] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 625–632.

[12] R. Min, A. Hadid, and J. L. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Proc. IEEE Conf. Autom. Face Gesture Recognit.*, Mar. 2011, pp. 442–447.

[13] H. Ekenel and R. Stiefelhagen, "Why is facial occlusion a challenging problem?" in *Advances in Biometrics* (Lecture Notes in Computer Science), M. Tistarelli and M. Nixon, Eds. Berlin, Germany: Springer-Verlag, 2009, pp. 299–308.

[14] J. Wright and Y. Ma, "Dense error correction via ℓ^1-minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

[15] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[16] C. X. Ren, D. Q. Dai, and H. Yan, "Coupled kernel embedding for low resolution face image recognition," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3770–3783, Aug. 2012.

[17] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Proc. Eur. Conf. Comput. Vis.*, 2010, pp. 448–461.

[18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Stat. Soc. B*, vol. 58, no. 1, pp. 267–288, 1996.

[19] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," Feb. 2012, to be published.

[20] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.

[21] B. A. Olshausen, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[22] F. Bach, "Consistency of the group lasso and multiple kernel learning," *J. Mach. Learn. Res.*, vol. 9, pp. 1179–1225, Jan. 2008.

[23] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[24] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Compressed sensing of block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[25] V. Cevher, M. Duarte, C. Hegde, and R. Baraniuk, "Sparse signal recovery using Markov random fields," in *Proc. Workshop Neural Inf. Process. Syst.*, Vancouver, BC, Canada, Dec. 2008, pp. 1–8.

[26] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[27] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[28] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts?" *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[29] X. T. Yuan and B. G. Hu, "Robust feature extraction via information theoretic learning," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 1193–1200.

[30] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[31] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[32] A. Martínez, "The AR face database," Comput. Vis. Center, Ohio State Univ., Columbus, Tech. Rep. 24, Jun. 1998.

Xiao-Xin Li received the B.E. degree in traffic civil engineering, the M.E. degree in computer application technology from the Wuhan University of Technology, Wuhan, China, in 2002 and 2005, respectively, and the Ph.D. degree in computer application technology from the South China University of Technology, Guangzhou, China, in 2009.

He is currently a Post-Doctoral Research Fellow with the Department of Mathematics, Faculty of Mathematics and Computing, Sun Yat-Sen University, Guangzhou. His current research interests include image processing, statistical data analysis, and bioinformatics.

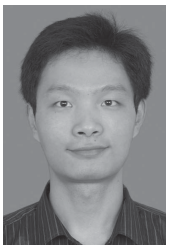
Dao-Qing Dai (M'09) received the B.Sc. degree from Hunan Normal University, Changsha, China, in 1983, the M.Sc. degree from Sun Yat-Sen University, Guangzhou, China, in 1986, and the Ph.D. degree from Wuhan University, Wuhan, China, in 1990, all in mathematics.

He was an Alexander von Humboldt Research Fellow with Free University, Berlin, Germany, from 1998 to 1999. He is currently a Professor and an Associate Dean of the Faculty of Mathematics and Computing, Sun Yat-Sen University. He has authored or co-authored over 100 refereed technical papers. His current research interests include image processing, wavelet analysis, face recognition, and bioinformatics.

Dr. Dai was the recipient of the Outstanding Research Achievements in Mathematics Award from the International Society for Analysis, Applications, and Computation, Fukuoka, Japan, in 1999. He served as a Program Co-chair of Sinobiometrics in 2004 and a Program Committee Member for several international conferences.

Xiao-Fei Zhang is currently pursuing the Ph.D. degree with the Department of Mathematics, Faculty of Mathematics and Computing, Sun Yat-Sen University, Guangzhou, China.

His current research interests include complex network, statistical data analysis, and bioinformatics.



Chuan-Xian Ren received the B.S. degree from Fuyang Normal University, Fuyang, China, in 2005, and the Ph.D. degree from the Mathematics Department, Faculty of Mathematics and Computing, Sun Yat-Sen (Zhongshan) University, Guangzhou, China, in 2010.

He was with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, in 2010 and 2011, as a Senior Research Associate. He is currently an Assistant Professor of the Faculty of Mathematics and Computing, Sun Yat-Sen University. His current research interests include image processing, face recognition, and statistical data analysis.