

# Transformed Principal Gradient Orientation for Robust and Precise Batch Face Alignment

Weihong Deng, Jiani Hu, Liu Liu, Jun Guo

Beijing University of Posts and Telecommunications, Beijing, China

**Abstract.** This paper addresses the problem of simultaneously aligning a batch of linearly correlated images despite large misalignment, severe illumination and occlusion. Our algorithm assumes that the gradient orientation of images, if correctly aligned, can be robustly represented by an underlying transformed principal gradient orientation (TPGO) subspace. With such a linear representation prior, the proposed method connects PGO subspace learning, gradient orientation reconstruction, and batch alignment in a unified framework with an efficient alternating optimization solution. Besides inherent robustness from the gradient orientation and the low-rank structure, TPGO maintains the pixel-accurate registration precision and the efficient optimization of Lucas & Kanade framework. Experimental results show TPGO based batch alignment is more precise and robust than the state-of-the-art methods such as RASL and SIFT feature base Congealing. Moreover, integrated with a SIFT based pre-alignment procedure, TPGO is able to align a large number of images of multiple objects with large deviation, illumination, and occlusion in the precision that surpasses the handcrafted alignments (provided by the standard database distributions), in term of our face recognition experiments on the Extended Yale B, AR and FERET databases.

## 1 Introduction

Batch Image alignment is an interesting task of automatic batch alignment of an ensemble of misaligned images to a fixed canonical template in an unsupervised manner. As the dramatic increase in the amount of visual data available, the image congealing technique has numerous applications in object detection, tracking, recognition, and retrieval. For instances, previous works on face recognition has proven that the recognition performance highly depend on the preciseness of the image alignment [1][2][3][4]. Two basic assumptions underlying most algorithms are that 1) images are subjected to randomly selected transformation of known nature, such as translation, similarity or affine, and that 2) images reside in a low dimensional subspace approximately after aligned to a fixed canonical template, e.g. faces [5][6], handwritten digits, etc. Clearly, the fundamental issue on batch image alignment is the multivariate similarity measures of an image ensemble, and the associated optimization methods.

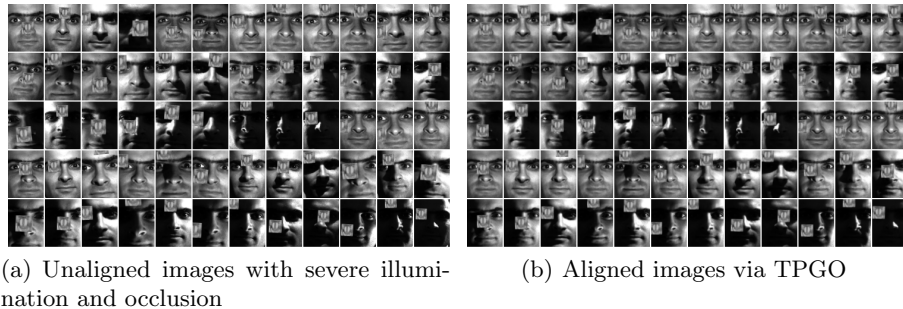
The pioneer work of Learned-Miller [7], named *congealing*, constructed the multivariate similarity measure as a sum of entropies of pixel values at each

pixel location in the whole image ensemble. To address the nonlinearity of entropy function, an sequential parameter update based optimization strategy is employed. In this congealing framework, clustered SIFT feature based entropy was also proposed to address the complex variations such as illumination and local deformation involved in the image ensemble[8]. To remedy the difficulty in the nonlinear optimization of congealing method, Cox et al. [9] proposed a least-square congealing method by employing the sum of squared distances between pairs of images to measure the similarity of the image ensemble. An inverse-compositional strategy was further proposed to address the “irrecoverable lost” problem caused by applying a single warp to a stack of images, which makes LS-Congealing applicable to align a large number of images. Both congealing methods ideally assume the matrix of aligned images will have exactly rank one, which may not be realistic for images with complex variations.

The other family of algorithms have been proposed to address the robust alignment problem by taking the advantage of the low-dimensional subspace structure. The early work of Frey and Jojic [10] used an EM algorithm to fit a low dimensional linear model, subject to domain transformation drawn from a know group. Schweitzer [11] proposed a more practical iterative procedure that jointly optimize the eigenspace model and the transformation parameters. Baker et al. used a similar technique to construct the active appearance model [12]. The Robust Parameterized Component Analysis (RPCA) algorithm used the robust fitting function to reduce the influence of occlusion and corruption. Vadaldi et al. [13] proposed a straightforward measure based on the dimension of subspace spanned by the aligned image, i.e. the rank of the image data matrix, but this measure may be sensitive to small corruption or occlusion of the images. To address this problem, Peng et al. [14] formulated a more robust measure by considering both the dimension of the subspace and the L1 norm of residuals from the subspace, and propose an influential method named RASL.

A major limitation of current batch alignment techniques is that their robustness is not enough to address the severe illumination and occlusion in the realistic images. Recent advance in image representation [15] have shown that it is indeed possible to invariantly represent facial images despite significant changes of illumination and occlusion, using the low-rank principal subspace of image gradient orientation. Inspired by this breakthrough work [15], this paper proposes a new algorithm to **achieve enhanced robustness by taking advantage of both the illumination/occlusion invariance and the low rank property of the aligned image gradient orientation**. Specifically, the contributions of this paper are as follows.

(1) **A new batch alignment algorithm called Transformed principal gradient orientation (TPGO)** is proposed for robustly aligning linear corrected images, despite uncontrolled lighting and large occlusions. Our algorithm assumes that the gradient orientation of unaligned image, if correctly aligned, can be robustly represented by a linear combination of the bases of an underlying low dimensional Transformed principal gradient orientation (TPGO) subspace. With such a linear representation prior, the proposed method connects PGO



**Fig. 1.** Batch Image Alignment via TPGO. (a) A full set of 60 images of the first person on the Extended Yale B database [16] with severe illuminations, simulated occlusions, and random transformations. For this challenging image ensemble, TPGO produces precise alignment within one pixel accuracy in the recovered eye centers, as shown in (b).

subspace learning, gradient orientation reconstruction, and batch alignment in a unified framework with an efficient alternative optimization solution. Besides inherent robustness from image gradient orientation [15], the “joint TPGO Subspace learning and batch Alignment” algorithm maintains the pixel-accurate registration precision (Fig. 1) and the efficient optimization of Lucas & Kanade framework [17], under the challenging variations of illuminations and occlusions. The effectiveness of the proposed TPGO is shown in term of the better alignment accuracy and robustness against two state-of-the-art batch alignment methods, namely SIFT-congealing [8] and RASL [14].

(2) **A coarse-to-fine batch alignment approach** is proposed to handle the task with large-misalignment and real-world illumination and occlusion. We observe that the real-world occlusions tend to largely deviate the bounding box of object detector, and makes the subsequent alignment algorithms cannot converge. Our approach applies a SIFT based pre-alignment procedure to coarsely align the images, before the fine alignment via TPGO. Experimental results on the 1000 images of 100 subjects from AR database show that our fully automatic batch alignment results (combined with VJ face detector) can reduce the recognition errors by a half, when compared to the manually aligned images distributed by the AR database.

(3) **TPGO is demonstrated be beneficial to fully automatic face recognition applications** where a TPGO subspace is first learned from the training images, and then the gallery images and probe images are aligned to the subspace so that they can be well compared. Recognition experiment on the FERET database with 1196 persons demonstrates that TPGO based alignment is more precise than SIFT-Congealing and Deformable sparse recovery method [18], as well as the manually labeled eye-coordinate based alignment, which is widely used as ground-truth alignment for this popular database.

## 2 Transformed Principal Gradient Orientation (TPGO)

In this section, we introduce a robust batch alignment approach named Transformed Principal Gradient Orientation (TPGO). The novelty of TPGO comes from the fact that it exploits both the illumination/occlusion-invariant descriptor and the low rank structure of aligned images, and, at the same time, maintains the preciseness of alignment and the elegance of the optimization.

### 2.1 Problem Formulation

**Image Gradient Orientations:** For an image of  $p$  pixels, we denote the image gradient at pixel  $i$  along the horizontal and vertical direction as  $g_{i,x}$  and  $g_{i,y}$  with  $i = 1, \dots, p$ . Hence, the gradient orientation (angle) can be computed by  $\phi_i = \arctan(g_{i,y}/g_{i,x}) \in [0, 2\pi)$ , where  $i = 1, \dots, p$ . In this way, the gradient orientation of each image  $I^i$  can be represented by a  $p$  dimensional complex vector  $\Phi^i = [\phi_1^i, \dots, \phi_p^i]^T \in \mathcal{R}^p$ . Alternatively, one can map the gradient orientation from an angle to a complex number  $\phi_k \rightarrow z_k = e^{j\phi_k}$ ,  $k = 1, \dots, p$ . In this sense, the gradient orientation of each image  $I^i$  can be represented by a  $p$  dimensional complex vector  $z^i = [e^{j\phi_1^i}, \dots, e^{j\phi_p^i}]^T \in \mathcal{C}^p$ .

**Gradient Orientation Distance:** For two images  $I^i$  and  $I^j$ , the local distance of the gradient orientations at pixel  $k$  is naturally defined as a cosine function of the angle difference, i.e.  $d^2(\phi_k^i, \phi_k^j) \triangleq [1 - \cos(\phi_k^i - \phi_k^j)]$ , which can be further formulated by the square distance between corresponding complex numbers, i.e.  $d^2(\phi_k^i, \phi_k^j) = \frac{1}{2} \|e^{j\phi_k^i} - e^{j\phi_k^j}\|^2$ . Therefore, the gradient orientation distance between two images  $I^i$  and  $I^j$ , the sum of the distances at each pixel, can be naturally formulated by the corresponding complex vectors, i.e.

$$d^2(\phi^i, \phi^j) = \frac{1}{2} \|z^i - z^j\|^2 \quad (1)$$

Besides its well-known invariance to illumination, the gradient orientation based distance is also robust to the occlusion of the images because the sum of distance computed from the occluded portion tends to be zero [15].

**Principal Gradient Orientation (PGO) Subspace:** Given a set of  $N$  images  $\{I^i\}_{i=1}^N$ , we compute the corresponding set of gradient orientation  $\{z^i\}_{i=1}^N$ . We look for a set of  $K < p$  orthonormal bases  $U = [u_1 \dots u_K] \in \mathcal{C}^{p \times K}$  with the goal of minimizing the sum of the distances from subspace

$$U^* = \arg \min_{U_K} \|Z - U_K U_K^H Z\|_F^2, \text{ s.t. } U^H U = I \quad (2)$$

where  $Z = [z^1 \dots z^N] \in \mathcal{C}^{p \times N}$ . The solution can be efficiently given by the  $K$  left singular vectors of  $Z$  corresponding to the  $K$  largest singular values. Recently, this subspace learning method has been successfully applied to the illumination- and occlusion-robust object recognition [15].

**Transformed PGO Subspace:** Given an ensemble of unaligned images, we assume that the gradient orientation of unaligned image, if correctly aligned, can

be robustly represented by a linear combination of the bases of an underlying low dimensional Transformed principal gradient orientation (TPGO) subspace. The term ‘‘Transformed’’ indicates that the underlying TPGO subspace would characterize the intrinsic structure of the aligned object that is invariant to the transformations of the specific images.

For the simplicity of formulation, we use  $z^i[\mathbf{p}^i]$  to denote the gradient orientation vector of the aligned image  $I^i(\mathbf{W}(\mathbf{x}; \mathbf{p}^i))$ ,  $T^i$  to denote the reconstructed template gradient vector of  $z^i[\mathbf{p}^i]$ , which is represented as a function  $T^i(U, \mathbf{p}^i)$  of the subspace  $U$  and transformation parameter  $\mathbf{p}^i$  itself. The objective function of TPGO is naturally formulated as follows.

$$E(U, \{T^i\}_{i=1}^N, \{\mathbf{p}^i\}_{i=1}^N) = \sum_{i=1}^N \|z^i[\mathbf{p}^i] - T^i(U, \mathbf{p}^i)\|^2 \quad (3)$$

## 2.2 Optimization Procedure

The proposed model (3) involves multiple variables and is hard to minimize directly. We adopt the alternating minimization scheme which reduces the original problem into several simpler subproblems. Specifically, we address the subproblems for each of the three variables in an alternating manner and present an overall efficient optimization problem. At each step, our algorithm reduces the objective function value, and finally converge to a local minima. To start, we initialize the alignment parameter  $\mathbf{p}^i = 0$ .

**PGO subspace estimation: Optimizing  $U_K$**  Given the current transformation parameter  $\mathbf{p}^i$  for each image, we want to update the bases of the underlying PGO subspace. We compute the corresponding set of gradient orientation  $\{z^i[\mathbf{p}^i]\}_{i=1}^N$ . We look for a set of  $K \ll p$  orthonormal bases  $U = [u_1 \cdots u_K] \in \mathcal{C}^{p \times K}$  with the goal of minimizing the sum of the distances from subspace

$$U^* = \arg \min_{U_K} \|Z[\mathbf{p}] - U_K U_K^H Z[\mathbf{p}]\|_F^2, \text{ s.t. } U^H U = I \quad (4)$$

where  $Z[\mathbf{p}] = [z^1[\mathbf{p}^1] \cdots z^N[\mathbf{p}^N]] \in \mathcal{C}^{p \times N}$ . The solution can be efficiently given by the  $K$  left singular vectors of  $Z[\mathbf{p}]$  corresponding to the  $K$  largest singular values.

**Latent Template Reconstruction: Optimizing  $T^i$**  With the underlying PGO subspace  $U_K$  and transformation parameter  $\mathbf{p}^i$ , we can reconstruct a latent template to be regarded as a virtual target for alignment. First, the current warped gradient orientation vector is projected onto the PGO subspace to obtain the reconstructed gradient as follows.

$$g^{i,t} = U_K U_K^H z^i[\mathbf{p}^i] \quad (5)$$

A normalization procedure is then applied to compute the template of gradient orientation

$$T_k^i = g_{k,x}^{i,t} / \|g_k^{i,t}\| + j g_{k,y}^{i,t} / \|g_k^{i,t}\| \quad (6)$$

**Image-to-Template Alignment: Optimizing  $\mathbf{p}^i$**  With the reconstructed template  $T^i$ , the minimization of the objective function (3) is decomposed to  $N$  subproblems on the maximization of the coherence between the warped gradient orientation vector and the corresponding template for each image, i.e.

$$\max_{\mathbf{p}^i} \sum_{k \in \mathbb{P}} z_{k,x}^i [\mathbf{p}^i] T_{k,x}^i + z_{k,y}^i [\mathbf{p}^i] T_{k,y}^i \quad (7)$$

In light of the inverse-compositional gradient correlation algorithm [19], the transformation parameters are updated by

$$\mathbf{W}(\mathbf{x}; \mathbf{p}^i) \leftarrow \mathbf{W}(\mathbf{x}; \mathbf{p}^i) \circ \mathbf{W}(\mathbf{x}; \Delta \mathbf{p}^i) \quad (8)$$

where  $\circ$  denotes composition, and

$$\Delta \mathbf{p}^i = \lambda (J^T J)^{-1} J^T S_{\Delta \phi} \quad (9)$$

where  $J$  is a  $p \times n$  Jacobian matrix whose  $k$ -th row has  $n$  element corresponding to the  $1 \times n$  vector

$$J_k = \frac{T_{k,x}^i \frac{\partial g_{k,y}^{i,t}}{\partial p} + T_{k,y}^i \frac{\partial g_{k,x}^{i,t}}{\partial p}}{\sqrt{(g_{k,x}^{i,t})^2 + (g_{k,y}^{i,t})^2}} \quad (10)$$

and

$$\begin{bmatrix} \frac{\partial g_{k,x}^{i,t}}{\partial p} \\ \frac{\partial g_{k,y}^{i,t}}{\partial p} \end{bmatrix} = \begin{bmatrix} g_{k,xx}^{i,t} & g_{k,xy}^{i,t} \\ g_{k,yx}^{i,t} & g_{k,yy}^{i,t} \end{bmatrix} \frac{\partial W}{\partial \mathbf{p}} \Big|_{\mathbf{p}=0} \quad (11)$$

$S_{\Delta \phi}$  is a  $N \times 1$  vector whose  $k$ -th element is equal to  $\sin(\phi_k^i[\mathbf{p}^i] - \phi_k^{i,t})$ .

### 2.3 Algorithm and Implementation Details

The overall algorithm optimizes the PGO bases  $U$ , latent gradient orientation template  $T^i$ , and alignment parameters  $\mathbf{p}^i$  alternatively. Algorithm 1 describes the procedures of our Joint TPGO Subspace learning and Batch Alignment algorithm

There are two loops in Algorithm 1. For each image, the inner loop aims to aligning it to the current subspace. In each inner loop, since the reconstructed template may not be accurate, we only update the transformation parameters once to avoid divergency. In contrast, the outer loop updates the subspace for more precise alignment.

### 2.4 Application to Fully Automatic Face Recognition

It should be noted that the proposed TPGO method is readily applicable to fully automatic face recognition. Specifically, In the training stage, algorithm 1 can be applied on a training image ensemble (or the gallery ensemble) to obtain a

**Algorithm 1** Joint TPGO Subspace learning and Batch Alignment algorithm

---

**Input:** An ensemble of unaligned image gradient orientation  $\{z^i\}_{i=1}^N$ ,  
**Output:** TPGO subspace bases  $U_K$ , transformation parameter  $\{\mathbf{p}^i\}_{i=1}^N$

- 1: **Initialization:** transformation parameter  $\mathbf{p} = 0$
- 2: **for**  $l_O = 1, 2, \dots, L_O$  **do**
- 3:   **Subspace Estimation:** update the PGO subspace  $U_K$  by minimizing Eqn.(4)
- 4:   **for**  $i = 1, 2, \dots, N$  **do**
- 5:     **for**  $l_I = 1, 2, \dots, L_I$  **do**
- 6:       **Template Reconstruction:** update the latent template  $T^i$  using Eqn. (6)
- 7:       **Image-to-Template Alignment:** update the transformation parameters  $\mathbf{p}^i$  (once) by (8)
- 8:     **end for**
- 9:   **end for**
- 10: **end for**

---

PGO subspace, which defines a fixed canonical template for image comparison. Then, in the test stage, the gallery image and probe image could be aligned to the PGO subspace so that they can be suitable compared. In this stage, as the PGO subspace is settle, the image can be aligned to the subspace by iteratively performing "template reconstruction" and "Image-to-Template Alignment" in algorithm 1.

Compared with the commonly used eye-coordinate based alignment, a advantage of this procedure is its full automation, which means no human labeling work involved in both training and testing stages. In addition, pixel-accurate alignment could be achieved by TPGO based alignment, but the error induced by eye localization is usually larger than one pixel, even for the human labeler. Therefore, it is possible that TPGO-based alignment could be better than eye-coordinate based alignment for recognition. We would test this possibility in our final experiment on the FERET database.

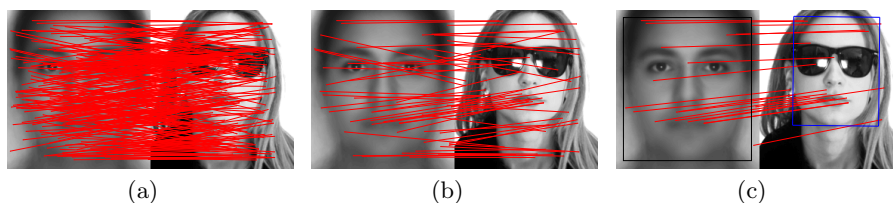
### 3 SIFT feature-based Generic Face Pre-alignment

In practice, severe illumination and occlusion not only affect the object appearance, but also largely deviate the bounding box of the object detector, which makes alignment algorithms more difficult to converge. For instance, we find that the scale of the bounding boxes of the commonly used Viola-Jones face detector is much larger for the face with sunglasses. These large initialized deviations are often outside the region of attraction for most batch alignment methods, especially when the images are with occlusion. To address this limitation, we proposed a SIFT feature based generic face pre-alignment approach<sup>1</sup>.

Inspired by the similar shape of the common face, our approach relies on a generic facial SIFT feature database which is constructed by extracting the

<sup>1</sup> Due to the space limit, the implementation details of the SIFT feature based pre-alignment is described in the supplementary material

SIFT feature points from a large set of aligned facial images. In our experiment, we have used 200 manually aligned faces from diverse sources. This generic feature database provides a prior distribution of the feature location of the human face. Given a novel input image, we first apply the Lowe’s matching algorithm to obtain a large number of matching point pairs between the input image and the generic database. Then, we eliminate a large proportion of the mismatching point pairs by adding a set of normalization constraints on the geometric information. Finally, based on the remaining matching point pair, we robustly estimate a similarity transformation (from the generic faces to the input face) by RANSAC algorithm. The implementation details is described in the supplementary material.



**Fig. 2.** (a) the feature match pairs using Lowe’s match algorithm, each red line connecting the left to right represents a estimated match point pairs, to simplify, we just show the location relationship of matched pairs without the scale and orientation of key points. We use a mean image to visualize the feature database from multiple images; (b) match point pairs after eliminating most of outliers by our normalization method (Step 2). (c) the outliers are further reduced by RANSAC, and the transformation from the black rectangle to the blue rectangle is the similarity transformation we calculate to roughly align the image.

## 4 Experiments

In this section, we demonstrate the efficacy of TPGO on batch alignment tasks despite severe lighting variation and occlusion, by comparing its performance with SIFT-Congealing and RASL. We test algorithms on a large number of realistic and challenging images taken from the Extended Yale B (EYB) database [16], the AR database [20], and FERET database [21]. EYB database contains full set of illumination images for human faces and AR database involves different lighting conditions and real-world large occlusions caused by accessories such as sunglasses and scarf. The FERET database contains a large number of subjects with diverse variations. Therefore, they are ideal for evaluating the robustness of batch alignment algorithms.

For comparison purpose, we also implemented three state-of-the-art methods: (1) SIFT-Congealing [8]: a robust alignment approach to complex images by minimizing the sum of entropies of the dense SIFT features; (2) RASL [14]:



**Table 1.** Mean error of the registered eye centers using different batch alignment algorithms under different initialized error. The notation  $\downarrow$  characterizes the proportion of the error reduced by batch alignment.

Init	1.92±0.88	2.80±1.17	4.03±1.25
SIFT-Congeaing	1.84±1.15 ( $\downarrow$ 4%)	1.78±1.78 ( $\downarrow$ 36%)	2.32±1.51 ( $\downarrow$ 42%)
RASL	1.21±1.67 ( $\downarrow$ 37%)	1.43±1.97 ( $\downarrow$ 49%)	1.53±1.89 ( $\downarrow$ 62%)
TPGO	0.68±0.44 ( $\downarrow$ 65%)	0.76±0.49 ( $\downarrow$ 73%)	1.39±0.82 ( $\downarrow$ 66%)

robust alignment via sparse and low-rank decomposition; ( For the first two algorithms, we preserve all the default settings of the publicly available codes). For TPGO, the subspace dimension is set to 5 for EYB, and 15 for AR and FERET. The number of outer iterations is set to 10. For the  $l_O$ -th outer iteration, the number of inner iterations is set to  $L_I = \max(5, 15 - l_O)$ .

#### 4.1 Aligning images with severe illuminations and occlusion

The experiment involves the full set of 60 images of the first person on the EYB database (See Fig. 1). First, all the images are first aligned by two (manually labelled) eye centers. Then, we perturbed the two points of eye centers using a Gaussian noise of standard deviation  $\sigma$ . Finally, using the similarity warp which the original and perturbed points defined, we generate the similarity distorted image. Specifically, we generate three sets of permuted images with  $\sigma=\{3,4,5\}$ , in which the mean of the eye mislocation<sup>2</sup> of the eye centers are 1.92, 2.80, and 4.03 pixels, respectively. The performance of batch alignment algorithm is evaluated by the mean, as well as the standard deviation, of the eye mislocation in the aligned images. If the mean error is reduced after batch alignment, the algorithm is considered convergent.

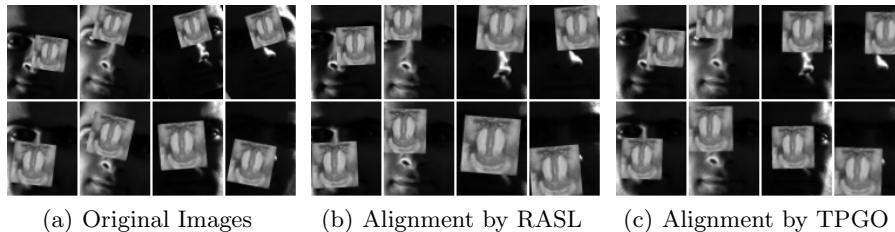
The comparative batch alignment performance of the three algorithms is enumerated in Table 1 and one can see from the Table that, in all three test cases, TPGO performs better than RASL, followed by SIFT-Congeaing. In addition, the standard deviation of the eye mislocation become larger after the batch alignment by RASL and SIFT-Congeaing. These results indicate that 1) local invariance of SIFT feature makes it not hard to perform pixel-accurate alignment; and 2) TPGO, which exploits both the illumination invariance and the low-rank structure of images, performs significantly stabler than RASL, which just considers the low-rank structure.

We further tests the robustness of the algorithms by adding synthetical occlusions to the illuminated images. On the three sets of images used in previous experiment, 10%, 20%, and 30% pixel occlusions are synthesized, respectively,

<sup>2</sup> In an image ensemble, mean location is calculated as the averaged coordinate of the (left or right) eye centers in all images. The mean of the eye mislocation is defined as the average value of all the distances between each eye center and its corresponding mean location.

**Table 2.** Mean error of the registered eye centers using different batch alignment algorithms under different initialized error and occlusion proportion. The notation  $\downarrow$  characterizes the proportion of the error reduced by batch alignment.

Init	$1.92 \pm 0.88$	$2.80 \pm 1.17$	$4.03 \pm 1.25$
10% Occlusion			
SIFT-Congealing	$1.90 \pm 1.11$ ( $\downarrow 1\%$ )	$2.14 \pm 1.63$ ( $\downarrow 24\%$ )	$2.65 \pm 2.10$ ( $\downarrow 34\%$ )
RASL	$1.61 \pm 2.30$ ( $\downarrow 16\%$ )	$1.68 \pm 1.77$ ( $\downarrow 40\%$ )	$1.97 \pm 2.22$ ( $\downarrow 51\%$ )
TPGO	$0.68 \pm 0.50$ ( $\downarrow 65\%$ )	$0.87 \pm 1.17$ ( $\downarrow 69\%$ )	$1.69 \pm 1.50$ ( $\downarrow 58\%$ )
20% Occlusion			
SIFT-Congealing	$2.27 \pm 1.52$ ( $\uparrow 18\%$ )	$2.60 \pm 2.16$ ( $\downarrow 7\%$ )	$3.66 \pm 2.70$ ( $\downarrow 9\%$ )
RASL	$1.62 \pm 1.67$ ( $\downarrow 16\%$ )	$2.15 \pm 2.16$ ( $\downarrow 23\%$ )	$2.45 \pm 2.58$ ( $\downarrow 39\%$ )
TPGO	$0.89 \pm 0.53$ ( $\downarrow 54\%$ )	$1.24 \pm 1.36$ ( $\downarrow 56\%$ )	$2.16 \pm 1.51$ ( $\downarrow 46\%$ )
30% Occlusion			
SIFT-Congealing	$3.11 \pm 1.91$ ( $\uparrow 62\%$ )	$3.17 \pm 2.14$ ( $\uparrow 13\%$ )	$3.72 \pm 2.36$ ( $\downarrow 8\%$ )
RASL	$2.24 \pm 2.07$ ( $\uparrow 17\%$ )	$2.58 \pm 2.41$ ( $\downarrow 8\%$ )	$2.99 \pm 2.43$ ( $\downarrow 26\%$ )
TPGO	$1.28 \pm 0.89$ ( $\downarrow 33\%$ )	$1.65 \pm 1.54$ ( $\downarrow 41\%$ )	$2.73 \pm 1.66$ ( $\downarrow 32\%$ )

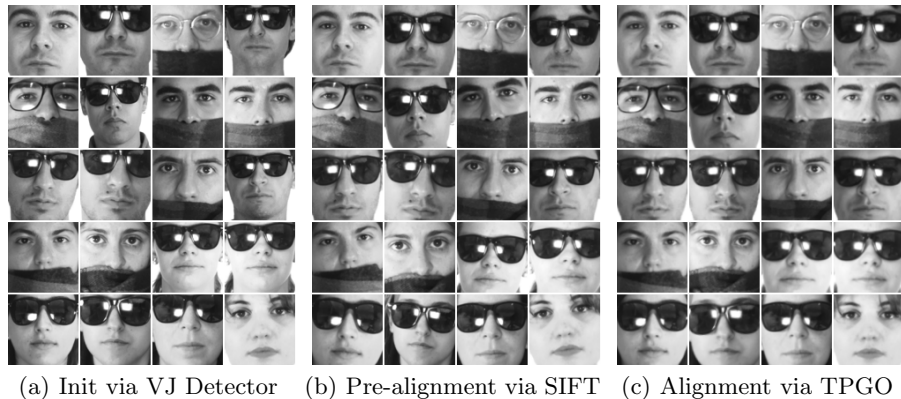


**Fig. 3.** The alignment results of very challenging images with large misalignment, severe illumination and occlusion.

before the image permutation. The results are listed in Table 2. As expected, all tested methods become worse when the occlusion proportion because larger. The mean and standard deviation of the proposed TPGO are smallest un all test cases. Under all nine test cases with occlusion, TPGO is the only method that can converge (reduce the mean error after batch alignment) all the time. Fig. 3 shows some alignment results that TPGO converges to reasonable results but RASL and SIFT-Congealing fails, as the illumination and occlusion become severer.

#### 4.2 Aligning 1000 images of 100 subjects with real-world illuminations and occlusions

This experiment involves a large number of facial images from the AR database. Unlike the synthetically occluded images in previous experiment, these images exhibit real-world large occlusions caused by sunglasses and scarves, in additional to lighting changes. Specifically, 1000 images of 100 subject from the Section



**Fig. 4.** Example images of the batch image alignment on the AR database.

1 of AR database are selected, with the conditions on natural light, left-side light, right-side light, both-side light, wearing sunglasses, sunglasses plus left-side light, sunglasses plus right-side light, wearing scarves, scarves plus left-side light, scarves plus right-side light, respectively (See Table 3 for examples).










We obtain an initial estimate of the transformation in each image using the VJ detector of OpenCV, followed by a SIFT feature based pre-alignment procedure<sup>3</sup>. After that, we align the images to an  $80 \times 80$  canonical frame using the three tested batch alignment methods. Finally, since there is no ground truth for this data set, we evaluate the preciseness of batch image alignment methods in term of the comparative recognition accuracy on their aligned image ensembles.

For each subject, the image with natural light is used as gallery, and the rest 9 images are used as probes. For the simplicity, whitened cosine similarity based nearest neighbor classifier is used for recognition. We have tested this classifier with LBP, Gabor, and HOG features, and find the LBP feature produce best results for all kinds of aligned images. Specifically, the  $LBP_{8,2}^{U_2}$  operator [23] is adopted in  $10 \times 10$  pixel cell, for each cell accumulating a local histogram of 59 uniform patterns over the pixels of the cell. The combined histogram entries form the representation, resulting in a 3776 ( $8 \times 8 \times 59$ ) dimensional feature vector.

Table 3 enumerates the comparative recognition accuracy on differently aligned ensembles. One can see from the table that (1) the cropped facial images via VJ detector receive a low recognition accuracy, especially for the images wearing sunglasses. This is because, as shown in Fig. 4(a), the sunglasses largely deviate the scale and translation of the bounding box, when compared with the non-occluded images. (2) Our pre-alignment procedure is effective to correct the

<sup>3</sup> None of the tested algorithms can align the images with sunglasses precisely without the SIFT based pre-alignment, because the initial transformation estimate of the off-the-shelf detector is largely biased by the occlusions. The detailed implementation of the pre-alignment procedure is described in the supplementary material.

**Table 3.** Evaluation of preciseness of batch image alignment methods in term of the comparative recognition accuracy (%) on their aligned image ensembles of the 1000 images from AR database. The natural image of each subject is used for template and the others are used as probes. The notation  $\downarrow$  characterizes the proportion of the recognition error reduced by the batch alignment method.

Alignment										Total
VJ-Detector	98	97	73	15	9	6	79	68	53	55.3
Pre-align	96	99	77	85	67	56	89	66	53	76.4
	$\uparrow 100\%$	$\downarrow 67\%$	$\downarrow 15\%$	$\downarrow 82\%$	$\downarrow 64\%$	$\downarrow 53\%$	$\downarrow 48\%$	$\uparrow 6\%$	$\downarrow 0\%$	$\downarrow 47\%$
Pre-align+	<b>100</b>	99	92	93	75	62	85	63	46	79.4
SIFT-Congeaing	$\downarrow 100\%$	$\downarrow 67\%$	$\downarrow 70\%$	$\downarrow 92\%$	$\downarrow 73\%$	$\downarrow 60\%$	$\downarrow 29\%$	$\uparrow 16\%$	$\uparrow 15\%$	$\downarrow 54\%$
Pre-align+	<b>100</b>	<b>100</b>	94	94	72	74	95	87	75	87.9
RASL	$\downarrow 100\%$	$\downarrow 100\%$	$\downarrow 78\%$	$\downarrow 93\%$	$\downarrow 69\%$	$\downarrow 72\%$	$\downarrow 76\%$	$\downarrow 59\%$	$\downarrow 47\%$	$\downarrow 73\%$
Pre-align+	<b>100</b>	<b>100</b>	<b>95</b>	<b>99</b>	<b>89</b>	<b>84</b>	<b>98</b>	<b>89</b>	<b>85</b>	<b>93.2</b>
TPGO	$\downarrow 100\%$	$\downarrow 100\%$	$\downarrow 81\%$	$\downarrow 99\%$	$\downarrow 88\%$	$\downarrow 83\%$	$\downarrow 90\%$	$\downarrow 66\%$	$\downarrow 68\%$	$\downarrow 85\%$
Handcrafts [22]	99	98	90	97	82	71	94	85	64	86.7

deviated detection of the occluded images, resulting in a notable improvement on recognition accuracy, as shown in 4(b). (3) All the three fine-alignment algorithms provide further improved accuracy based on the pre-alignment results. In particular, TPGO produces the highest accuracy on all the nine testing conditions. Some example of TPGO-aligned images are shown in 4(c).

Due to the difficulty in aligning the occluded images, AR database has provided a standard distribution of aligned images by the handcrafted approach described in [22]. To compare our automatic alignment with the manual alignment, the manually aligned images are first cropped to include similar facial region with our alignment, and then interpolated to the same size of  $80 \times 80$  pixels. As listed in Table 3, TPGO produces the higher accuracy than the handcrafted approach on all the nine testing conditions, and the overall error rate is reduced by over a half (from 13.5% to 6.8%). This result suggests that TPGO could potentially be very helpful for improving the performance of current object clustering or recognition systems despite large object occlusion.

**Speed and scalability of TPGO.** For this large-scale task, using 64-bit Matlab platform on a PC with Dual Core 2.93 GHz Pentium CPU and 4 GB memory, our implementation of TPGO requires less than 8 minutes to align the 1000 images of size  $80 \times 80$ , whereas RASL requires over 3 hours. This impressive computational efficiency is a direct result of using correlation of gradient orientation, instead of L1-norm of pixel intensity, for robust optimization.

### 4.3 Fully Automatic Face Recognition

In this section, we evaluate the effectiveness of TPGO on fully automatic face recognition using 3307 facial images of 1196 subjects from the gray-level FER-

ET database, which is a standard testbed for face recognition technologies [21]. The tested images display diversity across gender, ethnicity, and age, and were acquired without any restrictions imposed on expression, illumination and accessories (for examples). Specifically, the experiment follows the standard data partitions of the FERET database:

- *gallery training set* contains 1,196 images of 1,196 people.
- *fb probe set* contains 1,195 images taken with an alternative facial expression.
- *fc probe set* contains 194 images taken under different lighting conditions.
- *dup1 probe set* contains 722 images taken in a different time.
- *dup2 probe set* contains 234 images taken at least a year later, which is a subset of the dup1 set.

Our algorithm starts with facial images detected by the common face detectors. Viola and Jones face detector<sup>4</sup>, which outputs a square bounding box indicating the predicated center of the face and its scale, is applied for its stable performance and high speed. Given a detected face image of the width  $w$ , we crop the face according to the eye locations<sup>5</sup> of  $(0.305w, 0.385w)$  and  $(0.695w, 0.385w)$  using the CSU face identification evaluation system [24]. The cropped and scaled face images of a standard size  $150 \times 130$ , which subsequently is referred to as “*detected faces*”. These detected faces are used for the initialization of TP-GO learning. Since the detection deviation of the FERET images is small, the pre-alignment is not involved in this experiment.

It is well-known that sparse Representation-based Classification<sup>6</sup> (SRC) [25] is sensitive to the pixel-level misalignment of image, we therefore use it to evaluate the precision of alignment for automatic recognition. To solve the misalignment problem in SRC, a Deformable Sparse Recovery and Classification (DSRC) [18] have used tools from sparse representation to address the alignment problem. For the simplicity, TPGO and SIFT-Congeealing<sup>7</sup> first build an appearance model from the batch alignment of the gallery set, and than align the probe images to the model for recognition.

For comparison purpose, we also apply SRC to the eye-aligned faces and the detected faces. For a fair comparison, all the aligned faces are all downsampled to  $75 \times 65$  to be compatible with those used in [18]. The face recognition performance of SRC using the five alignment methods is tabulated in Table I, which shows that the best performance on three of the four probe sets is achieved

<sup>4</sup> We use the OpenCV implementation of the Viola and Jones’s face detector. Since there is only one face in each image, we reduce the false alarms by reserving the bounding box of the maximum size in each image. The detector missed only six faces out of all the 3307 images involved in our experiments, and we have manually completed these six bounding boxes.

<sup>5</sup> They are roughly the averaged locations of the two eyes of the typical bounding faces determined by the VJ face detector.

<sup>6</sup> The Homotopy method is applied to solve the  $\ell^1$ -minimization problem with the regularization parameter  $\lambda = 0.003$ .

<sup>7</sup> RASL has not been tested since it is not directly applicable to align unseen images for automatic recognition.

using TPGO-based faces. DSRC performs better than TPGO+SRC only when expression variation (fb set) is presented. In contrast, using TPGO-aligned faces achieves substantially improved accuracy (about 6% to 18%) than other alignment methods on the fc, dup1, and dup2 probe sets. This suggests that TPGO constructs a unified appearance model that is more robust against the complex variations of the facial appearance.

**Table 4.** Comparative FERET recognition rates on differently aligned faces using SRC

Alignment	fb	fc	dup1	dup2
Human labeled Eye-aligned faces+SRC	83.2	74.2	46.1	30.8
Detected faces+SRC	73.5	38.7	34.5	33.3
DSRC [18]	<b>95.2</b>	28.4	46.1	20.3
SIFT-Congeaing + SRC	82.0	73.2	55.3	42.3
TPGO-aligned faces+SRC	87.9	<b>82.4</b>	<b>61.2</b>	<b>50.0</b>

## 5 Conclusions

We have presented an image alignment method that can simultaneously align multiple images by exploiting both the illumination/occlusion invariance and the low rank property of the aligned image gradient orientation. Our approach is based on recent advances in image representation of gradient orientation that come with theoretical guarantees. The proposed algorithm consists of solving an efficient alternating optimization. This allows us to simultaneously align hundreds or even thousands of images on a typical PC in matter of minutes. We have shown the efficacy of our method with extensive experiments on images taken under laboratory conditions and on natural images of various types taken under a wide range of real-world conditions.

Experimental results show TPGO based batch alignment is more precise and robust than the state-of-the-art methods such as RASL and SIFT feature base Congeaing. Furthermore, integrated with our proposed SIFT based pre-alignment procedure, TPGO is able to align a large number of images of multiple objects with large deviation, illumination, and occlusion in the precision that surpasses the handcrafted alignment, in term of our face recognition experiments on the AR and FERET databases.

### Acknowledgement.

This work was partially sponsored by National Natural Science Foundation of China (NSFC) under Grant No. 61375031, No. 61471048, and No. 61273217. This work was also supported by the Fundamental Research Funds for the Central Universities, Beijing Higher Education Young Elite Teacher Project, and the Program for New Century Excellent Talents in University.

## References

1. Deng, W., Hu, J., Guo, J., Cai, W., Feng, D.: Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognition* **43**(5) (2010) 1748–1762
2. Deng, W., Hu, J., Lu, J., Guo, J.: Transform-invariant pca: A unified approach to fully automatic face alignment, representation, and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**(6) (2014) 1275–1284
3. Deng, W., Hu, J., Guo, J.: Extended src: Undersampled face recognition via intra-class variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(9) (2012) 1864–1870
4. Deng, W., Hu, J., Zhou, X., Guo, J.: Equidistant prototypes embedding for single sample based face recognition with generic learning and incremental learning. *Pattern Recognition* **47**(12) (2014) 3738–3749
5. Deng, W., Hu, J., Guo, J., Zhang, H., Zhang, C.: Comments on “globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics”. *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(8) (2008) 1503–1504
6. Deng, W., Liu, Y., Hu, J., Guo, J.: The small sample size problem of ica: A comparative study and analysis. *Pattern Recognition* **45**(12) (2012) 4438–4450
7. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. *Journal IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(2) (2006) 236–250
8. Huang, G.B., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, IEEE* (2007) 1–8
9. Cox, M., Sridharan, S., Lucey, S., Cohn, J.: Least squares congealing for unsupervised alignment of images. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE* (2008) 1–8
10. Frey, B.J., Jojic, N.: Transformed component analysis: Joint estimation of spatial transformations and image components. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. Volume 2., IEEE* (1999) 1190–1196
11. Schweitzer, H.: Optimal eigenfeature selection by optimal image registration. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. Volume 1., IEEE* (1999)
12. Baker, S., Matthews, I., Schneider, J.: Automatic construction of active appearance models as an image coding problem. *IEEE Trans. Pattern Anal. Machine Intell.* **26**(10) (2004) 1380–1384
13. Vedaldi, A., Guidi, G., Soatto, S.: Joint data alignment up to (lossy) transformations. In: *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE* (2008) 1–8
14. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE* (2010) 763–770
15. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **34**(12) (2012) 2454–2466

16. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence* **23**(6) (2001) 643–660
17. Baker, S., Matthews, I.: Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision* **56**(3) (2004) 221–255
18. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.: Toward a practical face recognition system: Robust alignment and illumination by sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **34**(2) (2012) 372–386
19. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Robust and efficient parametric face alignment. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*, IEEE (2011) 1847–1854
20. Martinez, A.M., Benavente, R.: The ar face database. CVC Technical Report #24 (June 1998)
21. Phillips, P.J., Moon, H., Rizvi, P., Rauss, P.: The feret evaluation method for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.* **22** (2000) 0162–8828
22. Martínez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **24**(6) (2002) 748–763
23. Ahonen, T., Hadid, A., Pietikinen, M.: Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(12) (2006) 2037–2041
24. Bolme, D.S., Beveridge, J.R., Teixeira, M., Draper, B.A.: The csu face identification evaluation system: its purpose, features, and structure. In: *Computer Vision Systems*. Springer (2003) 304–313
25. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust Face Recognition via Sparse Representation. *IEEE Trans. Pattern Anal. Machine Intell.* **31**(2) (2009) 210–227