Single-Sample Face Recognition with Image Corruption and Misalignment via Sparse Illumination Transfer *

Liansheng Zhuang[1,2], Allen Y. Yang[2], Zihan Zhou[3], S. Shankar Sastry[2], and Yi Ma[3]

[1]University of Science & Technology of China, Hefei, China, lszhuang@ustc.edu.cn
[2]Department of EECS, UC Berkeley, CA, {yang,sastry}@eecs.berkeley.edu
[3]Department of ECE, University of Illinois, Urbana, IL, {zzhou7,yima}@illinois.edu

Abstract

Single-sample face recognition is one of the most challenging problems in face recognition. We propose a novel face recognition algorithm to address this problem based on a sparse representation based classification (SRC) framework. The new algorithm is robust to image misalignment and pixel corruption, and is able to reduce required training images to one sample per class. To compensate the missing illumination information typically provided by multiple training images, a sparse illumination transfer (SIT) technique is introduced. The SIT algorithms seek additional illumination examples of face images from one or more additional subject classes, and form an illumination dictionary. By enforcing a sparse representation of the query image, the method can recover and transfer the pose and illumination information from the alignment stage to the recognition stage. Our extensive experiments have demonstrated that the new algorithms significantly outperform the existing algorithms in the single-sample regime and with less restrictions. In particular, the face alignment accuracy is comparable to that of the well-known Deformable SRC algorithm using multiple training images; and the face recognition accuracy exceeds those of the SRC and Extended SRC algorithms using hand labeled alignment initialization.

1. Introduction

Face recognition is one of the classical problems in computer vision. Given a natural image that may contain a human face, it has been known that the appearance of the face image can be easily affected by many image nuisances, including background illumination, pose, and facial corruption/disguise such as makeup, beard, and glasses. Hence, to develop a face recognition system whose performance can be comparable to or even exceed that of human vision, the computer system needs to address at least the following three closely related problems: First, it needs to effectively model the change of illumination on the human face. Second. it needs to align the pose of the face. Third, it needs to tolerance the corruption of facial features that leads to potential gross pixel error against the training images.

In the literature, many well-known solutions have been studied to tackle these problems [13, 32, 14, 9], although a complete review of the field is outside the scope of this paper. More recently, a new face recognition framework called *sparse-representation based classification* (SRC) was proposed [26], which can successfully address most of the above problems. The framework is built on a subspace illumination model characterizing the distribution of a corruption-free face image sample (stacked in vector form) under a fixed pose, one subspace model per subject class [2, 1]. When an unknown query image is jointly represented by all the subspace models, only a small subset of these subspace coefficients need to be nonzero, which would primarily correspond to the subspace model of the true subject. Therefore, by optimizing the sparsity of such an overcomplete linear representation, the dominant nonzero coefficients indicate the identity of the query image. In the case of image corruption, since the corruption typically only affects a sparse set of pixel values, one can concurrently optimize a sparse error term in the image space to compensate for the corrupted pixel values.

In practice, a face image may appear at any image location with random background. Hence, a face detection and registration step is typically first used to detect the face image. Most of the methods in face detection would learn

*The authors were supported in part by ARO 63092-MA-II, DARPA FA8650-11-1-7153, ONR N00014-09-1-0230, and NSF CCF09-64215, NSFC No. 60933013 and 61103134, Fundamental Research Funds for the Central Universities (WK210023002), and the Science Foundation for Outstanding Young Talent of Anhui Province (BJ2101020001).

a class of local image features/patches that are sensitive to the appearance of key facial features [27, 23, 17]. Using either an active shape model [5] or an active appearance model [4], the location of the face can be detected even when the expression of the face is not neutral or some facial features are occluded [21, 12]. However, using these face registration algorithms *alone* is not sufficient to align a query image to training images for SRC. The main reasons are two-fold: First, except for some fast detectors such as Viola-Jones [23], more sophisticated detectors are expensive to run and require learning prior distribution of the shape model from meticulously hand-labeled training images. More importantly, these detectors would register the pixel values of the query image with respect to the *average* shape model learned from all the training images, but they typically cannot align the pixel values of the query image to the training images for the purpose of recognition, as required in SRC.

Following the sparse representation framework in [26, 24], we propose a novel algorithm to effectively extend SRC for face alignment and recognition in the small sample set scenario. We observe that in addition to the well-understood image nuisances aforementioned, one of the remaining challenges in face recognition is indeed the small sample set problem. For instance, in many biometric, surveillance, and Internet applications, there may be only a few training examples per subject that are collected in the wild, and the subjects of interest may not be able to undergo an extended image collection session in a laboratory.[1]

Unfortunately, most of the existing SRC-based alignment and recognition algorithms would fail in such scenarios. For starters, the original SRC algorithm [26] assumes a plurality of training samples from each class must sufficiently span its illumination subspace. The algorithm would perform poorly in the single sample regime, as we will shown in our experiment later. In [24], in order to guarantee the training images contain sufficient illumination patterns, the test subjects must further go through a nontrivial passport-style image collection process in a dark room in order to be entered into the training database. More recently, another development in the SRC framework is simultaneous face alignment and recognition methods [28, 15, 30]. Nevertheless, these methods did not go beyond the basic assumption used in SRC and other prior art that the face illumination model is measured by a plurality of training samples for each class. Furthermore, as shown in [24], robust face alignment and recognition can be solved separately as a two-step process, as long as the recovered image transformation can be carried over from the alignment stage to the recognition stage. Therefore, simultaneous face alignment and recognition could make the already expensive sparse optimization problem even more difficult to solve.

[1]In this paper, we use Viola-Jones face detector to initialize the face image location. As a result, we do not consider scenarios where the face may contain a large 3D transformation or large expression change. These more severe conditions can be addressed in the face detection stage using more sophisticated face models as we mentioned above.

1.1. Contributions

Single-sample face alignment and recognition represents an important step towards practical face recognition solutions using images collected in the wild or on the Internet. We contend that the problem can be solved quite effectively by a simple yet elegant algorithm. The key observation is that one sample per class mainly deprives the algorithm of an illumination subspace model for each individual class. We show that a *sparse illumination transfer* (SIT) dictionary can be constructed to compensate the lack of the illumination information in the training set. Due to the fact that most human faces have similar shapes, only one subject is often sufficient to provide images of different illumination patterns, although adding more subjects may further improve the accuracy. The subject(s) for illumination transfer can be selected outside the set of training subjects for recognition. Finally, we show that the other image nuisances, including pose variation and image corruption, can be readily corrected by a single reference image of *arbitrary illumination condition* per class combined with the SIT dictionary. The SIT dictionary also does not need to know the information of any possible facial corruption for the algorithm to be robust. To the best of our knowledge, this work is the first to propose a solution to perform facial illumination compensation in the alignment stage and illumination and pose transfer in the recognition stage.

In terms of the algorithm complexity, the construction of the SIT dictionary is extremely simple when the illumination data of the SIT subject(s) are provided, and it does not necessarily involve any dictionary learning algorithm. The algorithm is also fast to execute in the alignment and recognition stages compared to the other SRC-type algorithms because a sparse optimization solver such as those in [29] is now faced with much smaller linear systems.

This paper bears resemblance to the work called Extended SRC [6], whereby an intraclass variant dictionary was similarly added to be a part of the SRC objective function for recognition. Our work differs from [6] in that the proposed SIT dictionary can be constructed from a selection of independent subject(s) only for the purpose of illumination transfer. As a result, the SIT dictionary is impartial to the training classes. Furthermore, by transferring both the pose and illumination from the alignment stage to the recognition stage, our algorithm can handle insufficient illumination and misalignment at the same time, and allows for the single reference images to have arbitrary illumination conditions. Finally, our algorithm is also robust to moderate amounts of image pixel corruption, even though we do not need to include any image corruption examples in the SIT

dictionary, while in [6] the intraclass variant dictionary uses both normal and corrupted face samples. We also compare our performance with [6] in Section 4.

2. Sparse Representation-based Classification

In this section, we first briefly review the SRC formulation and introduce the notation.

Assume a face image $\boldsymbol{b} \in \mathbb{R}^d$ in grayscale can be written in vector form by stacking its pixels. In the training stage, given L training subject classes, assume n_i well-aligned training images $A_i = [\boldsymbol{a}_{i,1}, \boldsymbol{a}_{i,2}, \cdots, \boldsymbol{a}_{i,n_i}] \in \mathbb{R}^{d \times n_i}$ of the same dimension as $\boldsymbol{b}$ are sampled for the i-th class under the frontal position and various illumination conditions. These training images are further aligned in terms of the coordinates of some salient facial features, e.g., eye corners and/or mouth corners. For brevity, the training images under such conditions are said to be in the *neutral position*. Furthermore, we do not consider facial expression change in this paper. Based on the illumination subspace assumption, if $\boldsymbol{b}$ belongs to the i-th class, then $\boldsymbol{b}$ lies in the low-dimensional subspace spanned by the training images in A_i, namely,

$$\boldsymbol{b} = A_i \boldsymbol{x}_i. \tag{1}$$

In the query stage, the query image $\boldsymbol{b}$ may contain an unknown 3D pose that is different from the neutral position. In image registration literature [18, 13, 24], an image transformation can be modeled in the image domain as $\tau \in T$, where T is a finite-dimensional group of transformations, such as translation, similarity transform, and homography. The goal of the alignment is to recover the transformation τ, such that an unwarped query image $\boldsymbol{b}_0$ of the same subject in the neutral position can be written as $\boldsymbol{b}_0 \doteq \boldsymbol{b} \circ \tau = A_i \boldsymbol{x}_i$.

In robust face alignment, the issue is often further exacerbated by the cascade of complex illumination patterns and moderate image pixel corruption and occlusion. In the SRC framework [26, 24], the combined effect of image misalignment and sparse corruption is modeled by

$$\hat{\tau}_i = \arg \min_{\boldsymbol{x}_i, \boldsymbol{e}, \tau_i} \|\boldsymbol{e}\|_1 \quad \text{subj. to} \quad \boldsymbol{b} \circ \tau_i = A_i \boldsymbol{x}_i + \boldsymbol{e}, \tag{2}$$

where the alignment is achieved on a per-class basis for each A_i, and $\boldsymbol{e} \in \mathbb{R}^d$ is the sparse alignment error as the objective function. After linearizing the nonlinear image transformation function τ, (2) can be solved iteratively by a standard ℓ_1-minimization solver. In [24], it was shown that the alignment based on (2) can tolerate translation shift up to 20% of the between-eye distance and up to 30° in-plane rotation, which is typically sufficient to compensate moderate misalignment caused by a good face detector.

Once the optimal transformation τ_i is recovered for each class i, the transformation is carried over to the recognition algorithm, where the training images in each A_i are transformed by τ_i^{-1} to align with the query image $\boldsymbol{b}$. Finally, a global sparse representation $\boldsymbol{x}$ with respect to the transformed training images is sought by solving the following sparse optimization problem:

$$\begin{array}{rcl} \boldsymbol{x}^* & = & \arg\min_{\boldsymbol{x},\boldsymbol{e}} \|\boldsymbol{x}\|_1 + \|\boldsymbol{e}\|_1. \\ \text{subj. to} & & \boldsymbol{b} = \left[A_1 \circ \tau_1^{-1}, \cdots, A_L \circ \tau_L^{-1}\right] \boldsymbol{x} + \boldsymbol{e} \end{array} \tag{3}$$

One can further show that when the correlation of the face samples in A is sufficiently tight in the high-dimensional image space, solving (3) via ℓ_1-minimization guarantees to recover both the sparse coefficients $\boldsymbol{x}$ and very dense (sparsity $\rho \nearrow 1$) randomly signed error $\boldsymbol{e}$ [25].

3. Sparse Illumination Transfer

3.1. Single-Sample Alignment

In this section, we first propose a novel face alignment algorithm that is effective even when a very small number of training images are provided per class. In the extreme case, we specifically consider the *single-sample face alignment problem* where only one training image $\boldsymbol{a}_i$ of *arbitrary illumination* is available from Class i. The same algorithm easily extends to the case when multiple training images are provided.

To mitigate the scarcity of the training images, something has to give to recover the missing illumination model under which the image appearance of a human face can be affected. Motivated by the idea of transfer learning [7, 20, 16], we stipulate that one can obtain the illumination information for both alignment *and* recognition from a set of additional subject classes, called the *illumination dictionary*. The additional face images have the same frontal pose as the training images, and can be collected offline and can be different from the query classes $A = [A_1, \cdots, A_L]$. In other words, no matter how scarce the training images of the query classes are, one can always obtain a potentially large set of additional face images of unrelated subjects who may have similar face shapes as the query subjects and may provide sufficient illumination examples.

The illumination dictionary for an additional class $L+1$ is defined as follows. Assume face images of sufficient illumination patterns $(\boldsymbol{a}_{L+1,1}, \boldsymbol{a}_{L+1,2}, \cdots, \boldsymbol{a}_{L+1,n}) \doteq (\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_n)$ are samples from the class, further assume all images in vector form are normalized to have unit length. Then the illumination dictionary by the $(L+1)$-th subject can be written as the difference of two face images of the same shape:

$$C_1 = [\boldsymbol{c}_2 - \boldsymbol{c}_1, \cdots, \boldsymbol{c}_n - \boldsymbol{c}_1]. \tag{4}$$

The multiplication of $C_1 \boldsymbol{y}$ by vector $\boldsymbol{y}$ can further generate more complex illumination patterns that involve multiple images in the columns of C_1.

We need to emphasize here that although the construction of C_1 in (4) is straightforward, by no means it is the

only way to obtain an illumination dictionary. In the literature, many other algorithms are well known, such as the quotient image [22, 19] and edge-preserving filters [3]. The focus of this paper is not on the illumination transfer function per se, but how its application on face images can enable single-sample alignment and recognition under the SRC framework. In addition, the illumination transfer shown later in (5) can be solved by efficient ℓ_1-minimization algorithms. Therefore, it has speed advantages compared to other more sophisticated methods. This approach was also used in [6] in the definition of the intraclass variant dictionary, but only for recognition. We will compare the performance of the two methods in Section 4.

Another issue with the illumination dictionary is that, if additional subject classes beyond $L+1$ are provided, one can continue to construct additional dictionaries $C = [C_1, C_2, \cdots]$. However, a somewhat unconventional observation we have discovered during our experiment is that if the first dictionary C_1 is carefully chosen, a single additional subject class is sufficient to achieve extremely good performance for face alignment and recognition. In Section 4, we will show that using a single illumination class, our alignment accuracy using only one reference image is comparable to that of [24] using multiple reference images, and the subsequent recognition accuracy further exceeds those using manual alignment results.

Clearly, this singular subject needs to have the facial appearance that is close to the "mean face," which has been used in face recognition to refer to the average appearance of faces over a population [2]. On the other hand, using those examples with abnormal facial features such as glasses and beard could easily reduce the performance. Without loss of generality, we assume $C = C_1$ in this paper. In Section 4.4, we will examine the efficacy of designing different illumination dictionaries with more subjects.

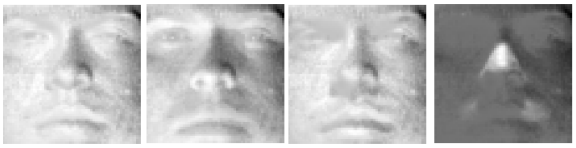

Figure 1. Examples of the elements of an illumination dictionary C constructed from the YaleB database.

Nevertheless, given the limited number of training images *in practice*, the illumination dictionary itself also cannot be arbitrarily large. Therefore, an effective solution should be able to achieve accurate alignment while only relying on a few illumination samples. Our solution is called *sparse illumination transfer* (SIT):

$$\begin{array}{rcl} \hat{\tau}_i & = & \arg\min_{\tau_i, x_i, \boldsymbol{y}_i, \boldsymbol{e}} \|\boldsymbol{y}_i\|_1 + \lambda\|\boldsymbol{e}\|_1, \\ \text{subj. to} & & \boldsymbol{b} \circ \tau_i = \boldsymbol{a}_i x_i + C\boldsymbol{y}_i + \boldsymbol{e} \end{array} \tag{5}$$

where λ is a parameter that balances the weight of $\boldsymbol{y}$ and $\boldsymbol{e}$, which can be chosen empirically. In our experiment, we found $\lambda = 1$ generally led to good performance for both uncorrupted and corrupted cases. Finally, the objective function (5) can be solved efficiently using ℓ_1-minimization techniques such as those discussed in [24, 29].[2] Figure 2 shows two examples of the alignment results.

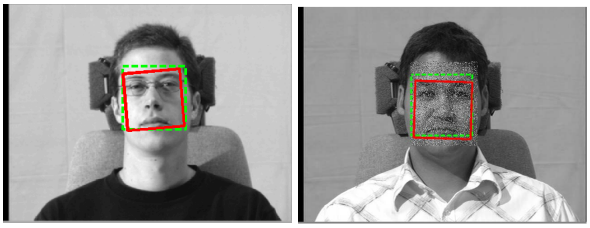

Figure 2. Single-sample alignment results on Multi-PIE. The solid red boxes are the initial face locations provided by a face detector. The dash green boxes show the alignment results. **Left:** The subject wears glasses. **Right:** The subject image has 30% of the face pixels corrupted by random noise.

3.2. Single-Sample Recognition

Next, we propose a novel face recognition algorithm that extends the SRC framework to the single-sample regime. Similar to the above alignment algorithm, the algorithm also applies trivially when multiple training samples per class are available.

Given the same reference image $\boldsymbol{a}_i$ as in (5), again we assume $\boldsymbol{a}_i$ is sampled from a random illumination condition. The key idea of our algorithm is to transfer and apply the estimated image transformation τ_i *and* the SIT compensation $C\boldsymbol{y}_i$ directly from the alignment step (5) to the recognition step. More specifically, for each reference image $\boldsymbol{a}_i$ of class i, define its warped version as

$$\tilde{\boldsymbol{a}}_i \doteq (\boldsymbol{a}_i x_i + C\boldsymbol{y}_i) \circ \tau_i^{-1}. \tag{6}$$

The modified reference image $\tilde{\boldsymbol{a}}_i$ aligns the orientation of $\boldsymbol{a}_i$ towards the query image, and at the same time adjusts the appearance of $\boldsymbol{a}_i$ to take into account the transferred illumination model $C\boldsymbol{y}_i$. Some examples about this effect are shown in Figure 3. After the SIT is applied to all the training images, we obtain the following *warped training dictionary* of L columns:

$$\tilde{A} = [\tilde{\boldsymbol{a}}_1, \cdots, \tilde{\boldsymbol{a}}_L]. \tag{7}$$

The SIT recognition algorithm solves a sparse representation of the query image $\boldsymbol{b}$ in the following linear system:

$$\begin{array}{rcl} \boldsymbol{x}^* & = & \arg\min_{\boldsymbol{x}, \boldsymbol{e}} \|\boldsymbol{x}\|_1 + \lambda\|\boldsymbol{e}\|_1, \\ \text{subj. to} & & \boldsymbol{b} = \tilde{A}\boldsymbol{x} + \boldsymbol{e} \end{array} \tag{8}$$

[2] In addition to seeking a sparse representation $\boldsymbol{y}$, an alternative solution could minimize the ℓ_2-norm of $\boldsymbol{y}$ instead, as used in [24, 31]. We have also tested the variation, and found the difference between the two solutions to the small, with minimizing $\|\boldsymbol{y}\|_1$ slightly better than minimizing $\|\boldsymbol{y}\|_2$.

Figure 3. Examples of warping a single reference image $\tilde{\boldsymbol{a}}_i = (\boldsymbol{a}_i + C\boldsymbol{y}_i) \circ \tau_i^{-1}$ for recognition. **Left:** Query image $\boldsymbol{b}$. **Middle Left:** Reference image $\boldsymbol{a}_i$. **Middle Right:** Illumination transfer information $C\boldsymbol{y}_i$. **Right:** Warped reference $\tilde{\boldsymbol{a}}_i$ has closer *pose* and *illumination* to $\boldsymbol{b}$ than the original image $\boldsymbol{a}_i$.

where the parameter λ can be chosen empirically.

In (8), the SIT dictionary $\tilde{A}$ only has L columns representing the training images from the L class, respectively. As a result, the recognition algorithm to recover the class label of $\boldsymbol{b}$ can be simplified from the original SRC algorithm [26], where the class corresponding to the largest coefficient magnitude in $\boldsymbol{x}$ is the estimated class of the query image $\boldsymbol{b}$. Figure 4 shows the estimated coefficients of an example of SIT recognition.

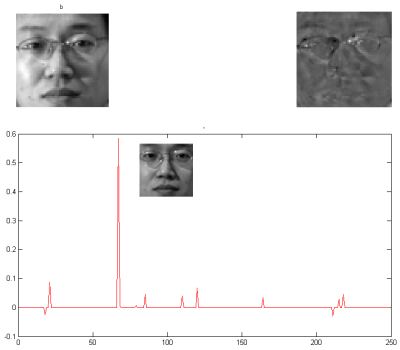

Figure 4. Illustration of SIT recognition. **Top Left:** $\boldsymbol{b}$. **Top Right:** $\boldsymbol{e}$. **Bottom:** Sparse representation $\boldsymbol{x}$ with the correct reference image $\boldsymbol{a}_i$ superimposed.

Before we move on to examine the performance of the new recognition algorithm (8), one may question the efficacy of enforcing a sparse representation in the constraint (8). The question may arise because in the original SRC framework, the data matrix $A = [A_1, \cdots, A_L]$ is a collection of highly correlated image samples that span the L illumination subspaces. Therefore, it makes sense to enforce a sparse representation as also validated by several followup studies [25, 8, 31]. However, in single-sample recognition, only one sample $\boldsymbol{a}_i$ is provided per class. Therefore, one would think that the best recognition performance can only be achieved by the nearest-neighbor algorithm.

There are at least two arguments to justify the use of sparse representation in (8). One one hand, as discussed in [26], in the case that $\boldsymbol{e}$ is a small dense error and the nearest-neighbor solution corresponds to a one-sparse binary vector $\boldsymbol{x}_0 = [\cdots, 0, 1, 0 \cdots]^T$ in the formulation (8), then solving (8) via ℓ_1-minimization can also recover the sparsest solution, namely, $\boldsymbol{x}^* \approx \boldsymbol{x}_0$. On the other hand, in the case that $\boldsymbol{e}$ represents a gross image corruption, as long as the elements of $\tilde{A}$ in (8) remain tightly correlated in the image space, the ℓ_1-minimization algorithm can compensate the dense error in the query image $\boldsymbol{b}$ [25]. This is a unique advantage over nearest-neighbor type algorithms.

4. Experiment

In this section, we present a comprehensive experiment to demonstrate the performance of our alignment and recognition algorithms. The illumination dictionary is constructed from YaleB face database [10]. YaleB contains 5760 single light source image of 10 subjects under 9 poses and 64 illumination conditions. For every subject in a particular pose, an image with ambient (background) illumination was also captured. In our experiments, we only use the first subject with its 65 aligned frontal images (64 illuminations + 1 ambient) to construct our illumination dictionary. The dictionary C is constructed by subtracting the ambient image from the other 64 illumination image. For a fair comparison, all the experiments in this section share the same YaleB illumination dictionary.

For the training and query subjects, we choose images from a much larger CMU Multi-PIE database [11]. Except for Section 4.3, 166 shared subject classes from Session 1 and Session 2 are selected for testing. In Session 1, we randomly select one frontal image per class with arbitrary illumination as the training image. Then we randomly select two different frontal images from Session 1 or Session 2 for testing. The outer eye corners of both training and query images are manually marked as the ground truth for registration. All the training face images are manually cropped into 60×60 pixels based on the locations of eyes out-corner points, and the distance between the two outer eye corners is normalized to be 50 pixels for each person. We again emphasize that our experimental setting is more practical than those used in some other publications, as we allow the training images to have arbitrary illumination and not necessarily just the ambient illumination.

We compare our algorithms with several state-of-the-art face alignment and recognition algorithms under the SRC framework. For the alignment benchmark, we compare with the deformable SRC (DSRC) algorithm [24] and the misalignment robust representation (MRR) algorithm [30].

For the recognition benchmark, we compare with DSRC, MRR based on the above automatic alignment results to find face regions. We also compare with the original SRC algorithm [26] and Extended SRC (ESRC) [6] with the face region location provided by manual labeling.

4.1. Simulation on 2D Alignment

We first demonstrate the performance of the SIT alignment algorithm dealing with simulated 2D deformation, including translation, rotation and scaling. The added deformation is introduced to the query images based on the ground truth coordinates of eye corners. The translation ranges from [-12, 12] pixels with a step of 2 pixels. Similar to [24], we use the estimated alignment error $\|\boldsymbol{e}\|_1$ as an indicator of success. More specifically, let $\boldsymbol{e}_0$ be the alignment error obtained by aligning a query image from the manually labeled position to the training images. We consider the alignment successful if $|\|\boldsymbol{e}\|_1 - \|\boldsymbol{e}_0\|_1| \leq 0.01\|\boldsymbol{e}_0\|_1$.

We compare our method with DSRC and MRR. As DSRC and MRR would require to have multiple reference images per class, to provide a fair comparison, we evaluate both algorithms under two settings: Firstly, seven reference images are provided per class to DSRC.[3] We denote this case as DSRC-7. Secondly, one randomly chosen image per class as the same setting as the SIT algorithm. We denote this case as DSRC-1 and MRR-1.

We draw the following observations from the alignment results shown in Figure 5:

1. SIT works well under a broad range of 2D deformation, particularly when the translation in x or y direction is less than 20% of the eye distance (10 pixels) and when the in-plane rotation is less than 30°.

2. Clearly, SIT outperforms both DSRC-1 and MRR-1 when the same setting is used, namely, one sample per class. The obvious reason is that DSRC and MRR were not designed to handle the single-sample alignment scenario.

3. SIT slightly outperforms DSRC-7, where DSRC-7 has access to seven training images of different illumination conditions. Furthermore, the SIT dictionary is derived from a single subject class from the unrelated YaleB database. It validates that illumination examples of a well-chosen subject are sufficient for SIT alignment.

4.2. Single-Sample Recognition

In this subsection, we evaluate the SIT recognition algorithm based on single reference images of the 166 subject classes shared in Multi-PIE Sessions 1 & 2. We compare its performance with SRC, ESRC, DSRC, and MRR.

First, we note that the new SIT framework and the existing sparse representation algorithms are *not* mutually exclusive. In particular, the illumination transfer (6) can be easily adopted by the other algorithms to improve the illumination condition of the training images, especially in the single-sample setting. In the first experiment, we demonstrate the improvement of SRC and ESRC with the illumination transfer. Since both algorithms do not address the alignment problem, manual labels of the face location are assumed to be the aligned face location. The comparison is presented in Table 1.

Table 1. Single-sample recognition accuracy via manual alignment.

Method	Session 1 (%)	Session 2 (%)
SRC_M	88.0	53.6
ESRC_M	89.6	56.6
SRC_M + SIT	91.6	59.0
ESRC_M + SIT	**93.6**	**59.3**

We observe that since the training images are selected from Session 1, there is no surprise that the recognition rates of those testing images also from Session 1 are significantly higher than those of Session 2. The comparison further shows adding the illumination transfer information to the SRC and ESRC algorithms meaningfully improves their performance by 3% – 4%.

Second, we compare DSRC, MRR, and SIT in the full pipeline of alignment plus recognition shown in Table 2.

Table 2. Single-sample alignment + recognition accuracy.

Method	Session 1 (%)	Session 2 (%)
DSRC	36.1	35.7
MRR	46.2	34.6
SIT	**79.9**	**65.7**

Compared with the past reported results of DSRC and MRR, their recognition accuracy decreases significantly when only one training image is available per class. It demonstrates that these algorithm were not designed to perform well in the single-sample regime. In both Session 1 and Session 2, SIT outperforms both algorithms by more the 30%. It is more interesting to compare the Session 2 recognition rates in Table 1 and Table 2, the more difficult and realistic experiment. SIT that relies on a SIT dictionary to automatically alignment the testing images achieves 65.7%, which is even higher than the ESRC rate of 59.3% with manual alignment.

4.3. Robustness under Random Corruption

In this subsection, we further compare the robustness of the SIT recognition algorithm to random pixel corruption.

[3] The training are illuminations {0,1,7,13,14,16,18} in Multi-PIE Session 1.

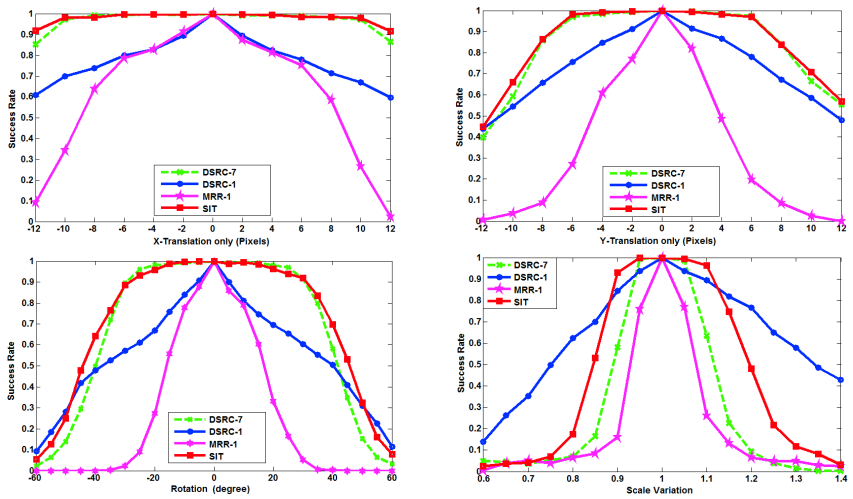

Figure 5. Success rate of face alignment under four types of 2D deformation: x-translation, y-translation, rotation, and scaling. The amount of translation is expressed in pixels, and the in-plane rotation is expressed in degrees.

We again compare the overall recognition rate of SIT with DSRC and MRR, the two most relevant algorithms.

To benchmark the recognition under different corruption percentage, it is important that the query images and the training images have close facial appearance, otherwise different facial features would also contribute to facial corruption or disguise, such as glasses, beard, or different hair styles. To limit this variability, in this experiment, we use Multi-PIE Session 1 for both training and testing, although the images should never overlap. We use all the subjects in Session 1 as the training and testing sets. For each subject, we randomly select one frontal image with arbitrary illumination for testing. Various levels of image corruption from 10% to 40% are randomly generated in the face region. Similar to the previous experiments, the face regions are detected by Viola-Jones detector. The performance of the three algorithms is shown in Table 3.

Table 3. Recognition rates (%) under various random corruption.

Corruption	10%	20%	30%	40%
DSRC	32.9%	31.7%	28.9%	24.1%
MRR	24.9%	14.5%	11.7%	9.2%
SIT	**74.3%**	**70.3%**	**67.1%**	**55.8%**

The comparison is more illustrative than Table 2. For instance, with 40% pixel corruption, SIT still maintains 56% accuracy; with 10% corruption, SIT outperforms DSRC and MRR by more than 40%.

4.4. Multiple-Subject SIT Dictionaries

The last topic of our discussion is the effect of choosing multiple subject classes for building the SIT dictionary, as we previously mentioned in (4). In the above alignment and recognition comparison, we have seen that SIT is comparable to or outperforms the existing face recognition algorithms using just a one-subject illumination dictionary. In this experiment, we provide some empirical observations to investigate the change of its alignment accuracy from using one subject to 10 subjects. Figure 6 again shows the alignment success rates when the face bounding box undergoes x-axis and y-axis translation, respectively, between [-12, 12] pixels.

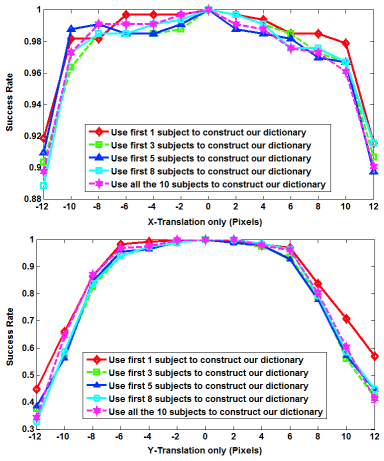

Figure 6. SIT alignment success rates from one to 10 subjects.

We observe that adjusting the size of the illumination dictionary does affect the alignment performance. However, the change is not monotonically increasing with more subject classes. In particular, for x-translation, all dictionaries are able to maintain good performance (above 98% recognition rate) even when the translation is as large as ± 10 pixels. For y-translation, the single-sample illumination dictionary slightly outperforms the others with more subjects when the translation is large.

5. Conclusion and Discussion

In this paper, we have presented a novel face recognition algorithm specifically designed for single-sample alignment and recognition. Although we have provided some exciting results that represent a meaningful step forward towards a real-world face recognition system, there remain several open problems that warrant further investigation. First, although the current way of constructing the illumination dictionary is efficient, the method is not able to separate the effect of surface albedo, shape, and illumination completely on face images. Therefore, a more sophisticated illumination transfer algorithm could lead to better overall performance. Second, although we have demonstrated empirically in Section 4.4 that including more subjects in the illumination dictionary may not necessarily lead to better performance, one could study whether a better dictionary learning algorithm could be applied to formulate the illumination dictionary that might represent more face shapes and illumination patterns.

References

[1] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *PAMI*, 25(2):218–233, 2003.

[2] P. Belhumeur, J. Hespanda, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *PAMI*, 19(7):711–720, 1997.

[3] X. Chen, M. Chen, X. Jin, and Q. Zhao. Face illumination transfer through edge-preserving filters. In *CVPR*, 2011.

[4] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *ECCV*, 1998.

[5] T. Cootes, C. Taylor, and J. Graham. Active shape models – their training and application. *CVIU*, 61:38–59, 1995.

[6] W. Deng, J. Hu, and J. Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *PAMI*, 34:1864–1870, 2012.

[7] C. Do and A. Ng. Transfer learning for text classification. In *NIPS*, 2005.

[8] E. Elhamifar and R. Vidal. Block-sparse recovery via convex optimization. *IEEE TSP*, 60:4094–4107, 2012.

[9] A. Ganesh, A. Wagner, J. Wright, A. Yang, Z. Zhou, and Y. Ma. Face recognition by sparse representation. In *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2011.

[10] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 23(6):643–660, 2001.

[11] R. Gross, I. Mathews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE FR*, 2008.

[12] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *ECCV*, 2008.

[13] G. Hager and P. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *PAMI*, 20(10):1025–1039, 1998.

[14] J. Ho, M. Yang, J. Lim, K. Lee, and D. Kriegman. Clustering appearances of objects under varying illumination conditions. In *CVPR*, pages 11–18, 2003.

[15] J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. In *CVPR*, 2008.

[16] C. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In *CVPR*, 2009.

[17] L. Liang, R. Xiao, F. Wen, and J. Sun. Face alignment via component-based discriminative search. In *ECCV*, 2008.

[18] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Int'l J. Conf. Artificial Intelligence*, 1981.

[19] P. Peers, N. Tamura, W. Matusik, and P. Debevec. Post-production facial performance relighting using reflectance transfer. In *SIGGRAPH*, 2007.

[20] A. Quattoni, M. Collins, and T. Darrell. Transfer learning for image classification with sparse prototype representations. In *CVPR*, 2008.

[21] J. Saragih, S. Lucey, and J. Cohn. Face alignment through subspace constrained mean-shifts. In *ICCV*, 2009.

[22] A. Shashua and T. Riklin-Raviv. The quotient image: Class-based re-rendering and recognition with varying illuminations. *PAMI*, 23(2):129–139, 2001.

[23] P. Viola and J. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004.

[24] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma. Toward a practical face recognition: Robust pose and illumination via sparse representation. *PAMI*, 34(2):372–386, 2012.

[25] J. Wright and Y. Ma. Dense error correction via ℓ^1-minimization. *IEEE TIT*, 56(7):3540–3560, 2010.

[26] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 31(2):210 – 227, 2009.

[27] S. Yan, C. Liu, S. Li, H. Zhang, H. Shum, and Q. Cheng. Face alignment using texture-constrained active shape models. *Image and Vision Computing*, 21:69–75, 2003.

[28] S. Yan, H. Wang, J. Liu, X. Tang, and T. Huang. Misalignment-robust face recognition. *IEEE TIP*, 19:1087–1096, 2010.

[29] A. Yang, Z. Zhou, A. Ganesh, S. Sastry, and Y. Ma. Fast ℓ_1-minimization algorithms for robust face recognition. Technical Report arXiv:1007.3753, University of California, Berkeley, 2012.

[30] M. Yang, L. Zhang, and D. Zhang. Efficient misalignment-robust representation for real-time face recognition. In *ECCV*, 2012.

[31] L. Zhang, M. Y. X. Feng, Y. Ma, and X. Feng. Collaborative representation based classification for face recognition. Technical report, arXiv:1204.2358, 2012.

[32] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comp. Sur.*, 2003.