Transformation-invariant Collaborative Sub-representation

Yeqing Li, Chen Chen, Jungzhou Huang
Department of Computer Science and Engineering
University of Texas at Arlington, Texas 76019, USA.
Email: yeqing.li@mavs.uta.edu, cchen@mavs.uta.edu, jzhuang@uta.edu

***Abstract*—In this paper, we present an efficient and robust image representation method that can handle misalignment, occlusion and big noises with lower computational cost. It is motivated by the sub-selection technique, which uses partial observations to efficiently approximate the original high dimensional problems. While it is very efficient, their method can not handle many real problems in practical applications, such as misalignment, occlusion and big noises. To this end, we propose a robust sub-representation method, which can effectively handle these problems with an efficient scheme. While its performance guarantee was theoretically proved, numerous experiments on practical applications have further demonstrated that the proposed method can lead to significant performance improvement in terms of speed and accuracy.**

I. Introduction

Image representation is an important problem in computer vision and pattern recognition and it has gained a lot of attentions in past decade. While huge interest has been seen in image representation, to date the most popular approaches are sparse representation and collaborative representation.

In Wright et al.'s pioneer work SRC [2] on sparse representation, they model the recognition problem as finding sparse representation of the test image based on linear combination of training images. Furthermore, outlier pixels are also assumed to be sparse. Favourable result has been achieved on face recognition application with occlusion and corruption. Zhang et al. proposed collaborative representation (CR) [3], [4] which relax the sparsity constraint on the representation coefficients. CR inspires many following works, such as MRR [5], DLRD_SR [6]. CR uses the l_2-norm regularization instead of l_1-norm in SRC. Therefore, it is much faster. It can achieve desired results for the clean data without corruption. However, if the testing images are sparsely corrupted, l_1-norm regularization has to be used for constraining sparse occlusion, which will lead to impressively degradation of speed. The high computation complexity is a big obstacle for them being used for high-dimensional images.

Moreover, in practice, due to the registration error of the object detector or the motion of the target object, the test image is usually not well-aligned with the training images. To achieve effective representation, the test image has to be aligned with training images first by using iterative transformation estimation methods [5], [7], [8]. This process has higher computation costs because the transformation estimation step is usually solved in high dimensional pixel-space and not able to utilize dimension reduction techniques. Though existing methods have achieved great success in various situations such as illumination change, occlusion and misalignment, their computational inefficiency limits them being used in practical application involving high-dimensional images.

To this end, we propose a robust sub-representation method, which can not only efficiently represent the image but also effectively handle the problems of misalignment, occlusion and big noises. Its performance guarantee can be theoretically proved. While combining it with existing collaborate representation method, a Transformation-invariant Collaborative Sub-Representation (TCSR) algorithm is proposed in this paper. Numerous experiments on practical applications have been conducted to further demonstrate its superior performance in terms of both computational complexity and accuracy.

The contributions of this paper are: **1)** To handle big noises, sub-selection method is generalized to robust sub-representation. We have theoretically proved its benefit over sub-selection method; **2)** Sub-representation is further extended to handle misalignment, occlusion and corrupted pixels, etc.. This extension is done by combing it with some existing techniques like transformation estimation algorithm and collaborative representation. The resulting method can also be regarded as a case study of sub-representation, which shows its great potential in cooperating with other methods. **3)** Extensive experiments are conducted to validate the efficiency and effectiveness of the proposed methods, which demonstrates that the proposed method can significantly accelerate the collaborative representation with imperceptible loss of accuracy.

II. Related Work

A. Sub-selection and Handling Incomplete Data

Sub-selection [9] is an efficient way to reduce the dimension of data. It projects the data onto lower dimensional subspaces using a randomly chosen subset of the data features. An example of sub-selection on pixel features of an image is shown in Fig. 1. In theoretical analysis, one situation that is similar to sub-selection is handling incomplete data. Balzano et. al. have recently proved theoretical guarantees of subspace detection using incomplete data [1].

The theory is sketched as follows for convenience. Let v_Ω be the vector of dimension $|\Omega| \times 1$ comprised of the elements v_i, $i \in \Omega$, ordered lexigraphically, where $|\Omega|$ denotes the cardinality of Ω. The energy of v in the subspace S is $\|P_S v\|_2^2$, where P_S denotes the projection operator onto S. Let U be an $n \times r$ matrix whose columns span the r-dimensional subspace S. In this case, $P_S = U(U^T U)^{-1} U^T$. With this representation in mind, let U_Ω denote the $|\Omega| \times r$ matrix, whose

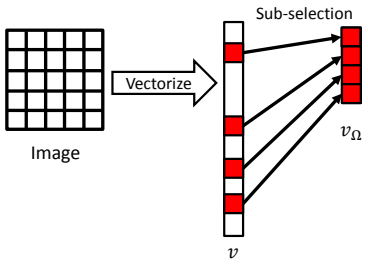


Fig. 1: Sub-selection on pixel features of an image.

rows are the $|\Omega|$ rows of U indexed by the set Ω, arranged in lexigraphic order. Suppose we only observe v on the set Ω. One approach for estimating its energy in S is to assess how well v_Ω can be represented in terms of the rows of U_Ω. Define the projection operator $P_{S_\Omega} := U_\Omega(U_\Omega^T U_\Omega)^\dagger U_\Omega^T$, where $\dagger$ denotes the pseudo-inverse. It follows immediately that if $v \in S$, then $\|v - P_S v\|_2^2 = 0$ and $\|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2 = 0$.

Let the entries of v be sampled uniformly with replacement. Again let Ω refer to the set of indices for observations of entries in v, and denote $|\Omega| = k$. The following theorem has been proved in [1]:

Theorem II.1. *(Theorem 1 in [1]):*

Let $\delta > 0$ and $k \geq \frac{8}{3} r\mu(S) log(\frac{2r}{\delta})$. Then with probability at least $1 - 4\delta$,

$$\frac{k}{n}(1 - \alpha - \alpha_0)\|v - P_S v\|_2^2 \leq \|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2 \leq (1 + \alpha)\frac{k}{n}\|v - P_S v\|_2^2 \quad (1)$$

where $\alpha = \sqrt{\frac{2\mu(y)^2}{k} log(\frac{1}{\delta})}$, $\beta = \sqrt{2\mu(y) log(\frac{1}{\delta})}$, $\gamma = \sqrt{\frac{8r\mu(S)}{3k} log(\frac{2r}{\delta})}$, and $\alpha_0 = \frac{r\mu(S)}{k} \frac{(1+\beta)^2}{1-\gamma}$.

$\mu(S)$ is the coherence of a subspace S [10]: $\mu(S) := \frac{n}{r} \max_j \|P_S e_j\|_2^2$, where e_j represents a standard basis element. Note that $1 \leq \mu(S) \leq \frac{n}{r}$. We let $\mu(y)$ denote the coherence of the subspace spanned by y. By plugging in the definition, we have $\mu(y) = \frac{n\|y\|_\infty^2}{\|y\|_2^2}$. Here, $v = x + y, x \in S$ and $y \in S^\perp$.

For general cases, Balzano et. al. have shown that if $|\Omega|$ is just slightly greater than $rlog(r)$, then with high probability $\|v_\Omega - P_{S_\Omega} v_\Omega\|_2^2$ is very close to $\frac{|\Omega|}{n}\|v - P_S v\|_2^2$. It means that partial observations can efficiently approximate the full observation with high dimension. Balzano et. al.'s result provides a useful starting point to analyse performance of sub-selection representation. However, it only concerns clean data and can not directly extend to handle many real problems in practical applications, such as the earlier noted misalignment, occlusion and big noises.

B. Transformation-invariant Collaborative Representation

Yang et al [5] proposed a transformation-invariant collaborative representation which can handle the representation and the transformation estimation simultaneously. First, a sparse term e is employed to handle occlusion. The problem is solved by minimizing the objective function:

$$\min_{x,e} F_1(x, e) = \|y - Ax - e\|_2^2 + \lambda\|x\|_2^2 + \gamma\|e\|_1, \quad (2)$$

where e represents the sparse big error, $y \in R^n$ is the query image, $A = [I_1, I_2, ..., I_p] \in R^{n \times p}$ is the dictionary, $x \in R^p$ denotes the representation coefficients. And then the transformation parameter is introduced so that the objective function becomes:

$$\min_{x,e,\tau} F_2(x, e, \tau) = \|y \odot \tau - Ax - e\|_2^2 + \lambda\|x\|_2^2 + \gamma\|e\|_1, \quad (3)$$

where τ is the transformation parameters and $y \odot \tau$ means to apply the transformation on the query image. The authors use a two-step strategy to accelerate the optimization process. Step one is to calculate SVD decomposition on the dictionary and use the singular vectors to solve x and τ:

$$\min_{\beta,e,\tau} F_3(\beta, e, \tau) = \|y \odot \tau - U\beta - e\|_2^2 + \gamma\|e\|_1, \quad (4)$$

where $A = USV^T$ is the SVD decomposition of A and $\beta = SV^T x$ is temporal representation coefficient with big absolute values. Minimizing this $F_3(\beta, e, \tau)$ will give an estimation of transformation parameters τ, big sparse error e, and the representation coefficients.

This approach is much faster than the previous sparse representation approach RASR [8]. However, due to the l_1-norm optimization and transformation estimation, it is still slower for high-dimensional images in practical applications.

III. Methodology

In this section, we introduce sub-representation approach and provide the theoretical proof of its performance guarantee for the case with sparse big noise. The resulting method is called as robust sub-representation. Finally, after combining it with transformation estimation and collaborative representation, we propose the TCSR approach.

In the following discussion, we shall interchangeably use τ to indicate the transformation parameters and an operator to apply that transformation on a set of images. Likewise, we may also use Ω as a sub-selection matrix as well as a sub-selection operation using that matrix. So that $y \odot \tau$ indicates applying the transformation on image y, and $A \odot \tau$ indicates applying the transformation on each image (column) in A. Similarly, $y \odot \Omega$ and $A \odot \Omega$ represents applying sub-selection on y and A respectively.

A. Sub-Representation

Consider the problem of representing a query image by linear combination of a set of images. Let $y \in R^n$ be the query image, $A = [I_1, I_2, ..., I_p] \in R^{n \times p}$ be the dictionary of p images, $x \in R^p$ denotes the representation coefficients. This problem can be formulated as solving the equation $Ax = y$. The dimension of this equation can be reduced using sub-selection that we discussed in Section II-A. Instead of solving $Ax = y$, we solve $A_\Omega \hat{x} = y_\Omega$. That is

$$\hat{x} = \underset{x}{\operatorname{argmin}} \|y_\Omega - A_\Omega x\|_2^2, \quad (5)$$

where $A_\Omega = A \odot \Omega$ denotes the dictionary under sub-selection, $y_\Omega = y \odot \Omega$ denotes the query image under sub-selection and $\hat{x}$ is the representation coefficient vector. From Theorem II.1, $\hat{x}$ should be very close to original x when Ω satisfy certain conditions. Equation 5 is our basic idea of sub-representation. The benefit of sub-representation is the solution of the original equation can be approximated by the solution of the low dimensional version of the equation. Hence, the the computational cost is significantly reduced. From now on, we shall extend sub-representation to handle several challenging problems and finally reach a practical image representation method.

B. Robust Sub-Representation

Occlusions and corrupted pixels are common challenges in many practical scenarios. In previous literatures, they are modelled as sparse big noise. Instead of solving $Ax = y$, we need to solve $Ax = y - e$, where e is the sparse error term. Adding sparse regularization on e, the problem becomes minimizing the following objective function:

$$\min_{x,e} F_4(x, e) = \gamma\|e\|_1 + \|y - Ax - e\|_2^2 \tag{6}$$

where γ is the regularization parameter. Solving this equation directly can be time consuming for even a medium size image, like 256×256, due to the l_1-norm minimization. However, usually $p \ll n$ or the rank of matrix A is far less than p and n, a sub-selection operation (e.g. $|\Omega| = m \ll n$) Ω can be applied on the representation. The new objective function is:

$$\min_{\hat{x},e_\Omega} G_4(\hat{x}, e_\Omega) = \gamma\|e_\Omega\|_1 + \|y_\Omega - A_\Omega\hat{x} - e_\Omega\|_2^2 \tag{7}$$

where $e_\Omega = e \odot \Omega$ is the error term resulting from acting sub-selection on e.

Now we shall discuss the relationship between Eq. (6) and Eq. (7), which is missing in the previous literatures. We first prove the boundedness of l_1-norm term under sub-selection. The follow theorem is required in later discussion:

Theorem III.1. *(McDiarmid's Inequality [11]): Let $X_1, ..., Xn$ be independent random variables, and assume f is a function for which there exist t_i, $i = 1, ..., n$ satisfying*

$$\sup_{x_1,...,x_n,\hat{x}_i} |f(x_1, ..., x_n) - f(x_1, ..., \hat{x}_i, ..., x_n)| \le t_i \tag{8}$$

where $\hat{x}_i$ indicates replacing the sample value x_i with any other of its possible values. Call $f(X_1, ..., X_n) := Y$. Then for any $\epsilon > 0$,

$$\mathbb{P}\left[Y \ge \mathbb{E}[Y] - \epsilon\right] \le \exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n t_i^2}\right) \tag{9}$$

$$\mathbb{P}\left[Y \le \mathbb{E}[Y] + \epsilon\right] \le exp\left(\frac{-2\epsilon^2}{\sum_{i=1}^n t_i^2}\right) \tag{10}$$

With Theorem III.1, the following lemma can be proved:

Lemma III.1. *: Suppose $\delta > 0$, $y \in \mathbf{R}^n$ and $|\Omega| = m$, then*

$$(1 - \alpha_1)\frac{m}{n}\|y\|_1 \le \|y_\Omega\|_1 \le (1 + \alpha_1)\frac{m}{n}\|y\|_1 \tag{11}$$

with probability at least $1 - 2\delta$.

Proof: We use McDiarmid's inequality from Theorem III.1 for the function $f(X_1, ..., X_m) = \sum_{i=1}^m X_i$ to prove this. Set $X_i = |y_{\Omega(i)}|$. Since $|y_{\Omega(i)}| \le \|y\|_\infty$ for all i, we have $\left|\sum_{i=1}^m X_i - \sum_{i\neq k} X_i - \hat{X_k}\right| = \left|X_k - \hat{X_k}\right| \le 2\|y\|_\infty$. We first calculate $\mathbb{E}[\sum_{i=1}^m X_i]$ as follows. Define $\mathbf{1}_{\{\}}$ to be the indicator function, and assume that the samples are taken uniformly with replacement. $\mathbb{E}\left[\sum_{i=1}^m X_i\right] = \mathbb{E}\left[\sum_{i=1}^m |y_{\Omega(i)}|\right] = \sum_{i=1}^m \left[\mathbb{E}\left[\sum_{j=1}^n |y_j|\mathbf{1}_{\{\Omega(i)=j\}}\right]\right] = \frac{m}{n}\|y\|_1$. Invoking the Theorem III.1, the left hand side is $\mathbb{P}\left[\sum_{i=1}^m X_i \le \mathbb{E}\left[\sum_{i=1}^m X_i\right] - \epsilon\right] = \mathbb{P}\left[\sum_{i=1}^m X_i \le \frac{m}{n}\|y\|_1 - \epsilon\right]$.

We can let $\epsilon = \alpha\frac{m}{n}\|y\|_1$ and then have that this probability is bounded by $\exp\left(\frac{-2\alpha^2(\frac{m}{n})^2\|y\|_1^2}{4m\|y\|_\infty^2}\right)$ Thus, the resulting probability bound is $\mathbb{P}\left[\|y_\Omega\|_1 \ge (1-\alpha)\frac{m}{n}\|y\|_1\right] \ge 1 - \exp\left(\frac{-\alpha^2 m\|y\|_1^2}{2n^2\|y\|_\infty^2}\right)$.

Substituting our definitions of $\mu_1(y) = \frac{n\|y\|_\infty}{\|y\|_1}$ and $\alpha_1 = \sqrt{\frac{2\mu_1(y)^2}{m} log(\frac{1}{\delta})}$ shows that the lower bound holds with probability at least $1 - \delta$. The argument for the upper bound can be proved similarly. The Lemma now follows by applying the union bound. ■

Lemma III.2. *: Let $\delta > 0$. Then with probability at least $1 - 6\delta$, $F_4(x, e)$ in Eq. (6) and $G_4(\hat{x}, e_\Omega)$ in Eq. (7) satisfy:*

$$\frac{m}{n}(1 - \alpha_4)F_4(x, e) \le G_4(\hat{x}, e_\Omega) \le \frac{m}{n}(1 + \alpha_4)F_4(x, e) \tag{12}$$

where α_4 is a small positive constant for given problem.

Lemma III.2 can be proved using Theorem II.1 and Lemma III.1. Due to limitation of pages, the detailed proof of this Lemma and later Lemmas are put in the supplementary material. With Lemma III.2, it's easy to derive that the sub-representation coefficient $\hat{x}$ will be very close to the origin solution x. Lemma III.2 provides the theoretical guarantee for quality of solution of Eq. (7).

C. Sub-selection and Transformation Estimation

Sub-selection can be used to accelerate transformation estimation. It can be formulated as:

$$\tau = \underset{\tau}{\operatorname{argmin}} \|y_1 - y_2 \odot \tau\|_2^2 \tag{13}$$

where y_1, y_2 are two images without correct alignment, τ is the unknown parameter of the transformation. For many kinds of transformation, like affine, similarity, homograph, translation, etc., the formula can be solved by an iterative approach [12]. To approximate Eq. (13), a Taylor expansion is usually applied. The problem becomes minimizing:

$$\min_{\Delta\tau} F_5(\Delta\tau) = \|y_1 - y_2 \odot \tau_c - J_\tau\Delta\tau\|_2^2, \tag{14}$$

where τ_c is the current estimation of the transformation, $\Delta\tau$ is the parameter increment at each iteration and J_τ is the Jacobian matrix. Eq. (14) is a least square equation and have close form solution. The dimension of the transformation parameters is usually very low compared to the image dimension. Hence, sub-selection is applicable here. Assume Ω is a sub-selection matrix(n choose m), $y_{\Omega,1} = y_1 \odot \Omega$, $y_{\Omega,2} = y_2 \odot \Omega$ and

$J_{\tau,\Omega} = Jacobian_{\hat{\tau}}(y_{\Omega,2})$. The result objective function is as follow:

$$\min_{\Delta\hat{\tau}} G_5(\Delta\hat{\tau}) = \|y_{\Omega,1} - y_{\Omega,2} \odot \hat{\tau}_c - J_{\hat{\tau},\Omega}\Delta\hat{\tau}\|_2^2. \quad (15)$$

Then, we have:

Lemma III.3. *: Let $\delta > 0$. Then with probability at least $1 - 4\delta$, $F_5(\Delta\tau)$ in Eq. (14) and $G_5(\Delta\hat{\tau})$ in Eq. (15) satisfy:*

$$\frac{m}{n}(1-\alpha_5)F_5(\Delta\tau) \leq G_5(\Delta\hat{\tau}) \leq \frac{m}{n}(1+\alpha_5)F_5(\Delta\tau)$$

where α_5 is a small positive constant for given problem.

Lemma III.3 can be proved using Theorem II.1. With this Lemma, $\Delta\hat{\tau}$ should be very close to $\Delta\tau$. This property allows us to address the unresolved problem of solving transformation estimation in the same low dimensional space of solving the representation coefficients. Lemma III.3 provides theoretical guarantee for quality of solution of Eq. (15).

D. Transformation Invariant Collaborative Sub-Representations (TCSR)

Now, we complete our discussion by combining techniques we discuss above and reach our final proposed method. The basic idea is transform Eq. (3) to low dimensional space via sub-selection. There resulting formula is as follow:

$$\min_{\hat{x},e_\Omega,\hat{\tau}} G_2(\hat{x}, e_\Omega, \hat{\tau}) = \|y_\Omega \odot \hat{\tau} - A_\Omega\hat{x} - e_\Omega\|_2^2 + \lambda\frac{m}{n}\|\hat{x}\|_2^2 + \gamma\|e_\Omega\|_1, \quad (16)$$

where $\hat{x}, \hat{\tau}$ are the representation parameters and transformation parameters respectively. Then we have:

Lemma III.4. *: Let $\delta > 0$*

Then with probability at least $1-6\delta$, $F_2(x, e, \tau)$ in Eq. (3) and $G_2(\hat{x}, e_\Omega, \hat{\tau})$ in Eq. (16) satisfy:

$$\frac{m}{n}(1-\alpha_7)F_2(x) \leq G_2(\hat{x}) \leq \frac{m}{n}(1+\alpha_7)F_2(x)$$

where α_7 is a small positive constant for given problem.

This Lemma can be proved using Lemma III.2, Lemma III.3 and Theorem II.1. Lemma III.4 provides theoretical guarantee for quality of solution of Eq. (16). Lemma III.4 is useful to prove the bound of sub-selection version of Eq. (4):

$$\{\hat{\beta}, e_\Omega, \hat{\tau}\} = \underset{\hat{\beta},e_\Omega,\hat{\tau}}{\operatorname{argmin}} \left\{ \|y_\Omega \odot \hat{\tau} - U_\Omega\hat{\beta} - e_\Omega\|_2^2 + \gamma\|e_\Omega\|_1 \right\} \quad (17)$$

where U_Ω is the singular vectors of sub-selection dictionary $\Omega \odot A = A_\Omega = \Omega \odot USV^T = U_\Omega SV^T$ and other terms have the same meaning as in the above discussion. The number of rows of Eq. (16) and Eq. (17) are much smaller than that of Eq. (3) and Eq. (4), which can effectively reduce the computational complexity. An object classification algorithm based on TCSR is summarized in Algorithm 1.

Algorithm 1 Transform-invariant Collaborative sub-representation (TCSR)

Input: Training data matrix A, query image y, and initial transformation τ_0 of y.
01: Generate l random selection operator $\Omega^1, ..., \Omega^l$
02: **for** $q = 1, ..., l$ **do**
03: Compute $A_\Omega = A \odot \Omega^q$
04: Compute $y_\Omega = y \odot \Omega^q$
05: Set $\hat{U}_1$ as the first η_1 column vectors of U_Ω where $A_\Omega = U_\Omega SV^T$
06: Solving Eq. (16) to get $\hat{x}, e_\Omega, \hat{\tau}^q$
07: Compute residue $r_i^q = \|y_\Omega \odot \hat{\tau}^q - A_\Omega\hat{x}_i\|$ for each class i
08: **end for**
09: Compute $\bar{r}_i = E[r_i]$
10: Compute $identity(y) = \text{argmin}_i\, E[r_i]$
11: Compute $transform(y) = E[\hat{\tau}^q]$
Output: $identity(y)$ and $transform(y)$

IV. Experiments

In this section, we conducted extensive experiments to validate the acceleration performance of the proposed sub-representation based methods against transformation estimation, collaborative representation, and transform-invariant collaborative representation, respectively. For fair comparisons, we download the code of these algorithms from their websites and follow their default parameter settings. Three databases are used for training or testing: Multi-PIE [13], Extended Yale B(EYB) [14] and AR [15]. All experiments are conducted on a desktop computer with Intel iCore 7 3.4GHz CPU. Matlab version is 2012a.

A. Transformation Estimation

First, we test our sub-selection for transformation estimation algorithm (STE, i.e. Eq. (15)) using public database CMU Multi-PIE database [13]. The images of 100 subjects from Session 2 are chosen and all images are resized to 640×480. The areas of human faces are used as the region of interest (ROI) and an artificial transformation of x and y directions are introduced to the ROI. The reference face area is resized to 160×120. We compare STE with the hierarchical model-based motion estimation (HMME, i.e. Eq. (14)) [12] algorithm. Artificial translation in both x and y directions are added to the position of ROI. Suppose the groundtruth rectangle position is a 4-D vector R_1 consisting of x, y coordinates, width and height of the rectangle, while R_2 is a similar vector for the result rectangle. The accuracy of the estimation is calculated as: $acc = \|R_1 - R_2\|_2/\|R_1\|_2$ The translation is set as 10, 20, 30, 40 pixels respectively. We set the sample rate as $1/5$ (i.e. $|\Omega| \approx n/5$). The result is shown in Table I. In the table, the proposed algorithm is 2 to 3 times faster than the original HMME algorithm while the loss of estimation accuracies is no more than 0.1%. These experiments indicate that the sub-selection technique effectively reduces the computational complexity while preserving the accuracies.

B. Face Recognition

We use the proposed TCSR (i.e. Algorithm 1) for face recognition to validate its benefits. We compare the proposed

Translation		10	20	30	40
HMME [12]	Acc.	**99.6%**	**99.1%**	94.9%	74.3%
	Time	0.19	0.38	0.60	0.43
proposed	Acc.	99.5%	**99.1%**	**95.5%**	**75.3%**
	Time	**0.07**	**0.15**	**0.21**	**0.14**

TABLE I: Result Comparison of Transformation Estimation. "Acc." stands for accuracy and "Time" stands for average execution time for each image.

TCSR with original collaborative representation classification(CRC) [4] algorithm on the Multi-PIE [13], Extended Yale B(EYB) [14] and AR databases [15]. For fair comparison, we follow the experimental setting in [4]. The initial position of all the testing and training images are automatically detected by Viola and Jone's face detector [16]. For the setting of Multi-PIE, the first 100 subjects in Session 1 with illuminations {0, 1, 7, 13, 14, 16, 18} are used as the training set while the first 100 subjects in Session 3 with illumination {3, 6, 11, 19} are used for testing. The face areas in the training images are all resized to 160×120. The sample rate is $1/15$; For the setting of EYB, 20 subjects are selected. For each subject, 32 randomly selected frontal images are used for training, with 29 of the remaining images for testing. The face areas are resized to 192×168. The sample rate is $1/10$; For the setting of AR, 100 subjects and 7 images per subject are selected as the training set, while 100 subjects and 6 images per subject for testing. There are 50 male faces and 50 female faces with various facial expressions and illuminations. The face areas are resized to 165×120. The sample rate is $1/8$;

		MPIE	EYB	AR
CRC [4]	Accuracy	88.7%	**99.4%**	86.1%
	Avg. Time	2.3	3.85	1.3
proposed	Accuracy	**89.1%**	99.1%	**86.6%**
	Avg. Time	**0.30**	**0.30**	**0.67**

TABLE II: Recognition rates and execution time

The experimental result is tabulated in Table II. The recognition rate of proposed TCSR algorithm is almost the same as the original CRC [4] algorithm with difference less than 0.5% while the speed is 3 to 10 times faster than the origin version related to the sample rate. Also the variation of illuminations of the training set and testing set affects the results. While the variation of the training set is sufficient to represent the query images, the sample rate can be lower, otherwise more data are needed. All these results have validated the proposed collaborative sub-representation has outperformed the collaborative representation in speed while achieving comparable accuracy.

C. Transformation Invariant Face Recognition

In this section, we conduct experiments to validate the benefit of TCSR under misalignment. First, we compare the proposed algorithm with state-of-the-art methods, MRR [5], RASR [8], TSR [7] on the Multi-PIE [13] database. The first 100 subjects in Session 1 with illuminations {0, 1, 7, 13, 14, 16, 18} are used as the training set while the first 100 subjects in Session 2 with illumination {10} are used for testing. Artificial transformation of 5 pixels translation in both x and y directions is added to the test images. The sample rate is set to $1/5$. The recognition rates of RASR, TSR, MRR

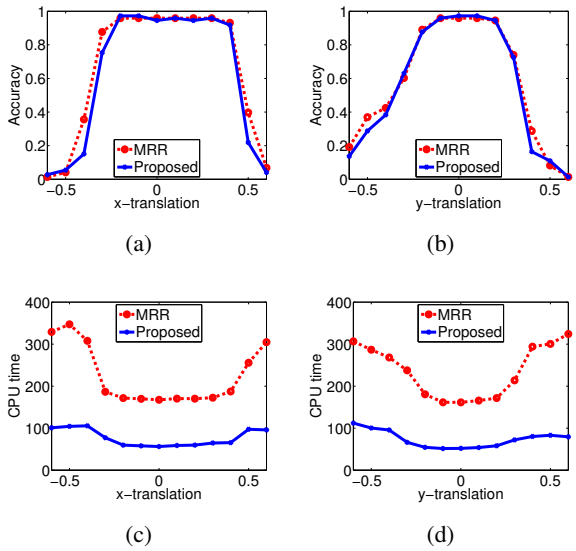


Fig. 2: Recognition rates and Speed with Translation: **Top row:** (a) Accuracy with only x direction translation; (b) Accuracy with only y direction translation; **Bottom row:** (c) Average Time with only x direction translation; (d) Average Time with only y direction translation;

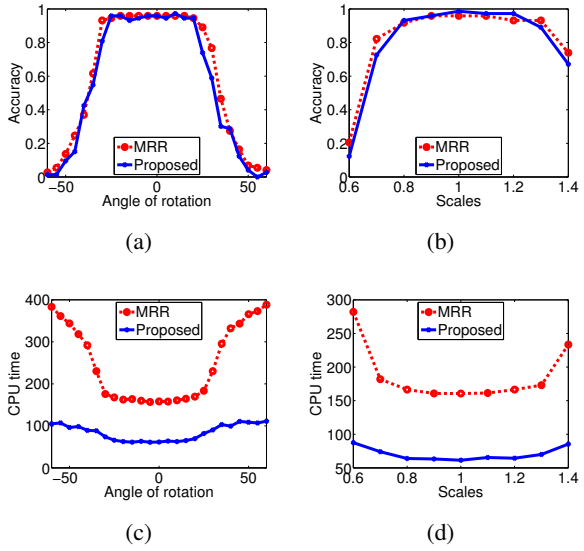


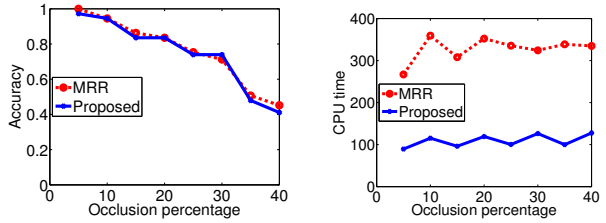
Fig. 3: Recognition rates and Speed with Scale and Rotation: **Top row:** (a) Accuracy with only in-place rotation; (b) Accuracy with only scale variantion; **Bottom row:** (c) Average Time with only in-place rotation; (d) Average Time with only scale variantion;

and the proposed algorithm are 91.8%, 89.1% 95.9% 95.9% respectively, while the average execution times are 95.9, 15.5, 4.9, 1.4 respectively. The RASR will try to fit all identities one by one, which makes it very slow in big training set. Although TSR is faster, it's less accurate. MRR is significantly faster than RASR and TSR and also more accurate. In the following

experiments we shall only compare our algorithm with MRR. Also note that in this experiment, our proposed algorithm is nearly 3 times faster than MRR while maintaining almost the same accuracy.

Next, we conduct experiments on the Multi-PIE database with various kinds of transformations. The experimental setting is the same as the previous experiment. Artificial transformations include x direction translation, y direction translation, in-place rotation and scales. These experiments compare the performance of MRR [5] and the proposed algorithm. Here we use the sample rate $1/5$. The experimental results are shown in Fig.2 and 3. The proposed algorithm is 3 to 4 times faster than the MRR, while the difference of their accuracy is less than 1%. Due to the redundancy of data, the sub-selection method preserve the accuracy very well. All these experiments validate that the proposed TCSR can handle various misalignments with much less computational cost than existing methods.

D. Recognition Despite Random Block Occlusion



(a) (b)

Fig. 4: Accuracy and CPU time (a) The recognition rate(accuray) vs. occlusion percentage; (b) The execution time vs. occlustion percentage

In this section, we further validate the robustness of TCSR by comparing our method and MRR [5] on images with random block occlusions. This kind of experiment has been conducted in the MRR and RASR [8], so we follow the setting of the experiment and use the dataset as the same as those used in the previous experiment. The sample rate is $1/3$. The training set and testing set are from the Multi-PIE database. The first 100 subjects in Session 1 are used as the training set. And the testing set is 100 subjects from Session 2. Various levels of block occlusion are added to the testing images. The testing results are shown in Fig.4. TCSR has almost the same recognition rates as those of the MRR under various levels of block occlusions. However, the TCSR is 2 to 3 times faster than the MRR algorithm. These experiments validate the benefit of proposed TCSR in handling occlusions.

V. Conclusion

This paper proposed a novel sub-representation method for image representation. We have theoretically proved its benefit for handling sparse big noise over previous methods. Combining it with existing techniques, we proposed a transform-invariant sub-representation, which can efficiently handle misalignment, occlusion and big noise problems in practical applications. The benefit of proposed methods were not only theoretically proved but also empirically validated by extensive experiment results on practical applications.

In the future, we would like to extend our approach in several directions. First, it is worth investigating beyond the simple sparse regularization to group sparsity [17]. Second, we would like to apply our approach to more applications such as MR imaging reconstruction problem [18], [19].

References

[1] L. Balzano, B. Recht, and R. Nowak, "High-dimensional matched subspace detection when data are missing," in *Proceedings of IEEE International Symposium on Information Theory*, 2010.

[2] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *PAMI 2009*, vol. 31, no. 2, pp. 210–227, 2009.

[3] R. Rigamonti, M. A. Brown, and V. Lepetit, "Are sparse representations really relevant for image classification?" in *CVPR 2011*. IEEE, 2011, pp. 1545–1552.

[4] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *ICCV 2011*. IEEE, 2011, pp. 471–478.

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

[6] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *CVPR 2012*. IEEE, 2012, pp. 2224–2231.

[7] J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," in *CVPR 2008*. IEEE, 2008, pp. 1–8.

[8] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *PAMI 2012*, vol. 34, no. 2, pp. 372–386, 2012.

[9] Y. Li, C. Chen, W. Liu, and J. Huang, "Sub-selective quantization for large-scale image search," in *Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2014.

[10] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[11] C. McDiarmid, "On method of bounded differences," *Surveys in Combinatorics*, vol. 141, pp. 148–188, 1989.

[12] J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani, "Hierarchical model-based motion estimation," in *ECCV'92*. Springer, 1992, pp. 237–252.

[13] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, no. 5, pp. 807–813, 2010.

[14] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *PAMI*, vol. 23, no. 6, pp. 643–660, 2001.

[15] A. Martınez and R. Benavente, "The ar face database," *Rapport technique*, vol. 24, 1998.

[16] P. Viola and M. J. Jones, "Robust real-time face detection," *IJCV*, vol. 57, no. 2, pp. 137–154, 2004.

[17] J. Huang, X. Huang, and D. Metaxas, "Learning with dynamic group sparsity," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 64–71.

[18] J. Huang, S. Zhang, H. Li, and D. Metaxas, "Composite splitting algorithms for convex optimization," *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1610–1622, 2011.

[19] J. Huang, S. Zhang, and D. Metaxas, "Efficient mr image reconstruction for compressed mr imaging," *Medical Image Analysis*, vol. 15, no. 5, pp. 670–679, 2011.