# Image-Set Matching by Two Dimensional Generalized Mutual Subspace Method


Bernardo B. Gatto
Institute of Computing (ICOMP)
Federal University of Amazonas, Brazil
bernardo@icomp.ufam.edu.br

Eulanda M. dos Santos
Institute of Computing (ICOMP)
Federal University of Amazonas, Brazil
emsantos@icomp.ufam.edu.br



***Abstract*—In this paper, we present a novel supervised learning algorithm for object recognition from sets of images, where the sets describe most of the variation in an object's appearance caused by lighting, pose and view angle. In this scenario, generalized mutual subspace method (gMSM) has attracted attention for image-set matching due to its advantages in accuracy and robustness. However, gMSM employs PCA, which has high computational cost contrasting to state-of-art appearance-based methods. To create a faster method, we replace the traditional PCA by 2D-PCA and variants on gMSM framework. In general, 2D-PCA and variants require less memory resource than conventional PCA since its covariance matrix is calculated directly from two-dimensional matrices. The introduced method has the advantage of representing the subspaces in a more compact manner, providing reasonably competitive recognition rate comparing to the traditional MSM, confirming the suitability of employing 2D-PCA and variants on gMSM framework. These results have been revealed through experimentation conducted on five widely used datasets.**


## I. Introduction

Classification and similarity between sets of images have been some of the crucial concerns in computer vision and image retrieval. Sets of images are able to represent the intrinsic variability of objects that have complicated shapes, producing more meaningful information for classification. Single-view image matching presents various drawbacks [1], such as: impossibility to recover a three-dimensional object from its two-dimensional single-view image (unless some restrictive assumptions about the real world are applied as prior knowledge), complex three-dimensional objects might cause self-occlusion and some features, necessary for recognition, might be lost. In this scenario, subspace-based approaches [2], have been employed in several computer vision applications, due to its considerable flexibility in dealing with multiple class problems and its straightforward implementation.

In subspace-based approaches, sets of images are expressed by its embedded subspace spanned by sets of basis vectors, where most of the variability of the image set is retained, requiring less memory for storage and speeding up the matching time. The subspace representation is achieved by principal component analysis (PCA) [3], which is optimal to achieve a subspace that minimizes the mean square error. This representation simplifies the classification of sets of images through the use of multiple canonical angles [4].

Following this main concept, generalized mutual subspace method (gMSM) [5] is a statistical pattern recognition method where each set of images is represented by a subspace and the similarity between these subspaces are determined by the use of multiple canonical angles. In gMSM, each subspace has a vector soft weigh, based on the eigenvalues. This approach differs from generalized difference subspace (GDS) [6], where only the eigenvectors with the highest eigenvalues are considered as basis vectors and the remainder are discarded. By employing weighed basis vectors to represent the subspaces, gMSM achieves high recognition performance.

Although gMSM has been employed successfully in the classification of sets of images, its performance is not satisfactory for more advanced systems, which more complicated structures should be classified, such as real-time systems, where memory requirements, algorithm complexity and execution time are critical components. Such systems are largely employed in augmented reality, where the applications would provide assistance to the users in medical procedures, educational systems and industrial design. In short, this issue is resulted from the fact that gMSM employs PCA in order to generate the subspaces as follows: First, each two-dimensional image from a set is reshaped to one-dimensional vector. Then, a covariance matrix is computed from these reshaped images. And finally, a set of basis vectors is generated from this covariance matrix and its weights are computed based on its eigenvalues. This reshaping procedure leads to a very high dimensional vector space, where the spatial structure of the reshaped images might be broken.

To overcome these drawbacks of gMSM, motivated by (2D-PCA) [7], we propose a two-dimensional generalized mutual subspace method (2D-gMSM) to speed up the learning and the matching processing times. The main difference between PCA and 2D-PCA is that 2D-PCA employs the image matrix directly, without vectoring the patterns, to generate the covariance matrix. Instead, in PCA, the matrix images are firstly vectorized, creating a covariance matrix which is larger than the covariance matrix produced by 2D-PCA. Since gMSM systematically operates on all the basis vectors produced by PCA, replacing PCA by 2D-PCA reduces the memory cost, as the basis vectors produced by 2D-PCA are more compact compared to the produced by PCA. As a consequence, 2D-gMSM is much more efficient than gMSM both in terms of memory complexity and time complexity. Therefore, 2D-gMSM presents three important improvements over conventional gMSM. First, 2D-gMSM operates on more meaningful subspaces, since the spatial structure of the images is preserved. Second, 2D-MSM is capable of handle RGB-D information, without the vectorization of the patterns, achieving more discriminative features. Third, the processing time to compute each subspace and memory requirements are reduced, due to the compactness of the subspaces achieved by 2D-PCA.

The organization of this work is as follows: Section 2 describe related work on image set classification and details the 2D-PCA and its variants. Then, in Section 3, we develop the two-dimensional generalized mutual subspace method for image set classification by introducing 2D-PCA and its variants on gMSM framework. Section 4 shows the advantages of the proposed method over the conventional gMSM by experimental results using ALOI [8] object dataset, RGB-D [9] object dataset, Honda/UCSD [10], YouTube Celebrities (YTC) [11] and PubFig83 [12] for face recognition. Finally, conclusions are discussed in Section 5.

## II. Related Work

In this section, we briefly describe 2D-PCA [7] and its variants: alternative 2D-PCA [13], Extended 2D-PCA [14], Color-PCA [15] and cross grouping 2D-PCA (C2D-PCA) [16]. This description is important in order to analyze the differences between traditional PCA and 2D-PCA variants. Although 2D-PCA have been developed a decade ago, recently several variants have been proposed, including cross grouping 2D-PCA [16]. In our paper, we investigate five 2D-PCA variants in order to identify the most suitable 2D-PCA variant that produces the best trade-off between accuracy and processing times. Therefore, it is important to analyze the impact of employing each variant.

### A. 2D-PCA and its Variants

In PCA, the theory states that the two-dimensional samples should be initially reshaped to one-dimensional vectors; otherwise, PCA cannot be employed. This reshaping process may break the structural information of the two-dimensional samples. In order to overcome this issue, 2D-PCA [7] was the first successful effort to apply PCA directly on two-dimensional images without reshaping the two-dimensional images into one-dimensional vectors. 2D-PCA inherits the same capabilities of the conventional PCA; however, its covariance matrix is calculated straightforwardly from two-dimensional matrices, instead of one-dimensional vectors. Therefore, the basis vectors achieved by 2D-PCA is much smaller than the basis vectors generated by the traditional PCA. In order to clarify this concept, let us consider $G_{2D}$ a 2D-PCA covariance matrix, which can be computed by:

$$G_{2D} = \frac{1}{M}\sum_{i=1}^{M}(A_i - A_\mu)^T(A_i - A_\mu), \tag{1}$$

where $\mathbf{A} = \{A_i\}_{i=1}^{M}$ is a set of two-dimensional images and $A_\mu$ is the mean image of $\mathbf{A}$. By eigen-decomposing $G_{2D}$, we obtain the optimal projection axes $\Phi_A = \{\phi_i\}_{i=1}^{M}$, which are the eigenvectors of $G_{2D}$ and have the following characteristics: $(i)$ $(\phi_i, \phi_j) = 0$, for any $i \neq j$ and $(ii)$ $(\phi_i, \phi_j) = 1$, for any $i = j$, where $(\cdot, \cdot)$ denotes inner product. In addition, the $\Phi_A$ set is ordered so that the first few $\phi_i$ vectors retain most of the variation available in the entire $\mathbf{A}$ set. In the case of data compression, the first $k$ vectors of $\Phi_A$, where $k \ll M$, represents most of the variation presents in $\mathbf{A}$. We can observe that by applying 2D-PCA instead of PCA we achieve a much more compact subspace. Therefore, we adopt the use of 2D-PCA to create the basis vectors instead of the traditional PCA, as will be detailed further.

Although 2D-PCA generates a more compact set of basis vectors than the conventional PCA, the feature extracted from a two-dimensional image is still a vector, not a matrix. Hence, the operation of extracting the basis vectors employed by 2D-PCA works systematically only in the row direction. An alternative denominated Alternative 2D-PCA [13] explores this idea and, instead of operating in the row direction, extracts the basis vectors by operating in the column direction. This approach showed that both column direction 2D-PCA and row direction 2D-PCA achieved similar performance, even working on orthogonal directions.

In order to explore within-row and between-row information of the covariance matrix, extended two-dimensional principal component analysis (E2D-PCA) [14] was proposed. In E2D-PCA, it is shown that the covariance matrix of 2D-PCA corresponds to the average of the main diagonal of the covariance matrix of PCA. Therefore, the covariance matrix achieved by 2D-PCA is a subset of the covariance matrix obtained by PCA. This subset may have less discriminative information than the original set, leading to a weaker discriminative set of features. Instead, E2D-PCA is able to produce more discriminative features than 2D-PCA by employing more covariance diagonals than the matrix covariance of 2D-PCA. Also, it is possible to directly control the trade-offs between recognition accuracy and model complexity.

In general, object recognition algorithms make use of gray-scale images for evaluating its performance. However, in [17] it is shown that color information plays an important role in face recognition systems. An extension of 2D-PCA denominated Color-PCA [15] was proposed in order to handle color information for face recognition systems. In addition, to explore the properties of 2D-PCA, Color-PCA also includes features of color images by maintaining RGB information as a third-order tensor. The higher recognition performance, compared to conventional 2D-PCA, is justified by the reason that the skin pixels would occur in close proximity to other skin pixels and that the skin color features would lie on a better discriminative subspace, which does not occur in gray-scale images. The procedure to create the principal components is similar to the 2D-PCA, except that the covariance matrix of each image is generated by concatenating the RGB color layers into a single matrix, instead of using just one gray-scale layer. In this case, the set of concatenated RGB layers images $\mathbf{A}^{RGB} = \{A_i^R \| A_i^G \| A_i^B\}_{i=1}^{M}$ is used in order to create the following correlation matrices:

$$C_H = \frac{1}{M}\sum_{i=1}^{M}(A_i^{RGB} - A_\mu{}^{RGB})^T(A_i^{RGB} - A_\mu{}^{RGB}), \quad (2)$$

$$C_V = \frac{1}{M}\sum_{i=1}^{M}(A_i^{RGB} - A_\mu{}^{RGB})(A_i^{RGB} - A_\mu{}^{RGB})^T, \quad (3)$$

where $C_H$ and $C_V$ stand for the correlation matrices when the images of $\mathbf{A}^{RGB}$ are concatenated horizontally and vertically.

2D-PCA and its variants use the 2D image matrices to construct the covariance matrix, grouping these features randomly by row or column of the input image. Thus, some informative patterns may be lost. To solve this issue, cross grouping 2D-PCA (C2D-PCA) [16] is proposed to face recognition. This technique aims to reduce the redundancy among the row and the column vectors of the image matrix. C2D-PCA completely preserves the covariance information of PCA between local geometric structures in the image matrix which is partially maintained in 2D-PCA and its variants. To accomplish these properties, the covariance matrix of C2D-PCA is produced from the summation of the outer products of the column and the row vectors of all images, then eigenvalue decomposition is applied to the covariance matrix in order to obtain the basis vectors employed to generate the subspaces.

### *B. Image Set classification Methods*

Several solutions to image set matching have been proposed in recent years. In general terms, these methods can be divided into two approaches: parametric model methods and non-parametric sample methods. The parametric model methods employ some parametric distribution, such as Gaussian, to describe each image set and then measure the distribution similarity. The non-parametric methods aim to describe an image set as a subspace or a manifold. These methods employ the distance between the manifolds or subspaces in order to measure the similarity between the image sets.

Discriminant Analysis of Canonical Correlations (DCC) [18] is an image set classification technique that attempts to find a subspace which the within-class correlation of sets is maximized and the between-class correlation is minimized. DCC uses a linear discriminative function to maximize canonical correlations of within-class sets and minimize canonical correlations of between-class sets. In DCC, the similarity of any two transformed data sets are defined as the sum of canonical correlations.

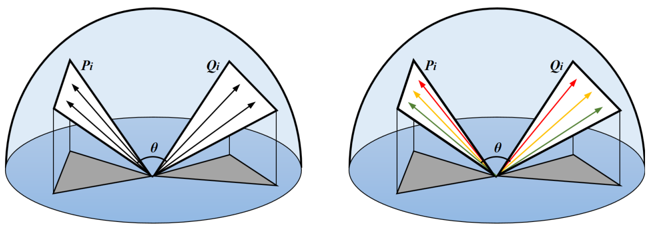
ManifoldManifold Distance (MMD) [19] represents each image set as a manifold. Each manifold consists of a collection of local linear subspaces, which can preserve large variations, such as illumination and point of view. The distances between pair-wise subspaces is integrated in order to create the similarity between the manifolds.

Manifold Discriminant Analysis (MDA) [20] is a manifold based image set classification technique that maximizes the distance of manifolds with different class labels and enhances the local data compactness within each manifold. MDA employs discriminative learning based on Linear Discriminant Analysis (LDA) in order to map the multi-class manifolds into an embedding space.

Convex Hull based Image Set Distance (CHISD) [21] is an image set classification technique that models each image set as a convex geometric region in feature space. The similarity between the convex geometric regions represented by convex hulls is computed based on the distance of closest point approach. By using a convex approximation, the method is less overfitting than the methods based on sample points because CHISD can produce new samples on the hull. In addition, the approach can be optimized to deal with outliers.

## III. Proposed Method

In this section, we first describe the steps to generate two dimensional subspaces by 2D-PCA and its variants. Then, we introduce the procedure to compute the weights of each two dimensional subspaces and the similarity based canonical angles. After that, we introduce the procedure of our object recognition framework. Finally, we show the computational advantage of replacing the conventional PCA by 2D-PCA and variants on gMSM. We conducted complexity analysis of 2D-PCA [7], A2D-PCA [13], E2D-PCA [14], Color-PCA [15] and cross grouping 2D-PCA (C2D-PCA) [16]. Figure 1 shows the difference between MSM and the proposed 2D-gMSM.




(a) (b)

Fig. 1. (a) The concept of MSM, where the subspace dimensions of $P_i$ and $Q_i$ are empirically obtained. (b) The proposed 2D-gMSM, where soft weighting evaluates the importance of each eigenvector. Therefore, 2D-gMSM employs all the basis vectors produced by 2D-PCA. Also, the basis vectors produced by 2D-PCA and its variants are more compact, improving the processing time.

### *A. Generating Subspaces by 2D-PCA*

To solve the image-set problem, we introduce 2D-gMSM, which is based on gMSM and 2D-PCA and variants. gMSM exploits the fact that a set of images lies in a cluster, which can be efficiently represented by a set of orthonormal basis vectors [2]. This approach is also applied by eigenspace [22]; however, in contrast to eigenspace, gMSM constructs a subspace for each different set of images, instead of just one. Our proposed method makes use of gMSM framework and, by replacing PCA by 2D-PCA, achieves a more compact and meaningful subspace. In our method, we represent a set of $M$ two-dimensional images $Y = \{Y_1, Y_2, \ldots, Y_M\}$ as a subspace, which is generated by extracting the eigenvectors from the following correlation matrix:

$$G_{2D} = \frac{1}{M}\sum_{i=1}^{M}(Y_i)^T(Y_i). \quad (4)$$

It should be noted that, in 2D-gMSM, the covariance matrix is not centered, different from the covariance matrix handled by 2D-PCA. In this work, we investigate the behavior of the proposed framework by replacing the covariance matrix in accordance with the 2D-PCA and its variants, creating a 2D-gMSM version for each 2D-PCA variant.

### *B. Computing the soft weighs of each Subspace*

As mentioned before, the basis vectors generated by 2D-PCA and its variants represent a set of images in a compact manner. In gMSM, all the eigenvectors are employed to represent a subspace. However, each eigenvector has its own weight, which is computed as follows; let $\Lambda_{2D} = diag(\lambda)$ be the eigenvalues of matrix $G_{2D}$ in descending order, the design of the soft weights is performed according to these eigenvalues. Let $\Omega = diag(w)$ be a diagonal matrix of soft weights:

$$\omega = w_M(\lambda) = \min\left[\frac{\lambda}{\lambda_M}, 1\right], \quad (5)$$

where $w_M$ is the $M-$th eigenvalue in $\lambda$. This soft weighting evaluates the importance of each eigenvector as a basis in the subspace by the variance relative to $\lambda_M$. The $M$ first values of the diagonal matrix $\Omega$ will be unity and the remainder will be proportionally decreasing with the $M-$th eigenvalue.

### *C. Similarity-based Canonical Angles*

After obtaining a set of basis vectors which best approximates each subspace to its corresponding set of images and its weights, we can compute the similarity between the subspaces. This procedure is achieved by applying canonical angles or principal angles [4]. As in gMSM, we consider that if the distance between two subspaces is smaller enough, then we consider these subspaces similar to each other. In practical terms, let $\Phi_{2D} = \{\phi_1, \phi_2, \ldots, \phi_M\}$ and $\Psi_{2D} = \{\psi_1, \psi_2, \ldots, \psi_M\}$ span two $M-$dimensional subspaces and $S_{\Phi_{2D},\Psi_{2D}} = \{0 \leq \theta_1 \leq \theta_2 \leq \ldots \leq \theta_n \leq \pi/2\}$ represents the set of angles between $\Phi_{2D}$ and $\Psi_{2D}$. A practical approach to determine $S_{\Phi_{2D},\Psi_{2D}}$ is by calculating the $\Lambda_{2D} = \{\lambda_1, \lambda_2, \ldots, \lambda_M\}$ eigenvalue of $\Omega_{\Phi_{2D}}\Phi_{2D}^T\Psi_{2D}\Omega_{\Psi_{2D}}$. The canonical angles $\theta_i = \{\cos^{-1}\lambda_1, \cos^{-1}\lambda_2, \ldots, \cos^{-1}\lambda_M\}$ are used to compute the structural similarity between soft weighted $\Phi_{2D}$ and $\Psi_{2D}$ subspaces as follows:

$$S(\Phi_{2D}, \Psi_{2D})_M = \frac{1}{M}\sum_{i=1}^{M}\cos^2\theta_i, \quad (6)$$

the structural similarities between subspaces are more robust to noise, such as illumination variations and point-of-view.

### *D. Image-set matching by using 2D-gMSM*

Let us assume that $C$ sets of training images are given by $\{A_1, A_2, \ldots, A_C\}$, where $A_i$ is a set containing $M$ two-dimensional images. Let us assume that each $A_i$ set belongs to one of the $C$ object classes. Then, we assume that there is a linear transformation that represents each $A_i$ set in terms of its variance. This new representation, $\{\Phi_1, \Phi_2, \ldots, \Phi_C\}$, provides a more compact manner to represent each $A_i$ set and its computational matching cost is therefore, greatly reduced. Each $\Phi_i$ basis vectors spans a reference subspace $P_i$, where its weights are computed based on its eigenvalues according to Eq. (5). Finally, for a given set of two-dimensional test images $Y = \{Y_1, Y_2, \ldots, Y_M\}$, the task is to compute a subspace $Q_Y$ that represents $Y$ in terms of its variance and predicts its corresponding image set based on the nearest $P_i$ reference subspace, according to Eq. (6).

### *E. Computational Advantage*

The main difference of 2D-gMSM from traditional gMSM is that 2D-gMSM does not require transforming image matrices into vectors. Thus, it reduces the computational complexity of construction of the subspaces and reduces the computation time of the matching. All these make the proposed algorithm superior to gMSM, in terms of computational time. In addition, the basis vectors produced by each 2D-PCA variant specifies the time required and the computational complexity of this algorithm. In 2D-PCA and its variants approach, the time requirements and the computational complexity in all these methods are close to each other. However, excepting from Color-PCA and E2D-PCA, all are smaller than PCA.

The components for constructing gMSM and 2D-gMSM are similar. In order to clarify, we adopt the computational advantage of 2D-gMSM over gMSM, since calculating the covariance matrix of 2D-PCA, A2D-PCA, and C2D-PCA, hold the same computational complexity [16]. However, Color-PCA requires more computational resource, since Color-PCA works on the RGB channels.

In order to show the computational advantage of 2D-gMSM over gMSM, let us follow the steps to extract both gMSM and 2D-gMSM basis vectors from the set of $M$ images $Y = \{Y_1, Y_2, \ldots, Y_M\}$. In gMSM, each $y_i$ image is previously reshaped to $d-$dimensional vectors, where $d = h \times w$. Then, let us denote $G_{gMSM}$ and $G_{2D-gMSM}$ as covariance matrices employed by gMSM and 2D-gMSM respectively.

In this scenario, it is required $2(d \times d \times M)$ flops (taking into account float point multiplications and additions) to compute both $G_{gMSM}$ and $G_{2D-gMSM}$ covariance matrices (see Eq. (4)). From the above, we obtain that the size of $G_{gMSM}$ and $G_{2D-gMSM}$ are respectively $d \times d$ and $h \times h$. The next step is the eigen-decomposition of $G_{gMSM}$ and $G_{2D-gMSM}$. The computational complexity of eigen-decomposing an $n \times n$ matrix is $O(n^3)$. Therefore, extracting the basis vectors from $G_{2D-gMSM}$ is computationally more efficient than extracting the basis vectors from $G_{gMSM}$, since $G_{2D-gMSM}$ is much smaller than $G_{gMSM}$, as well as the matching times.

The relationship between 2D-PCA and PCA is that the scatter matrix of 2D-PCA is constructed by sum of all scatter matrices of different column indices in the main diagonal of PCA [14]. Therefore, using 2D-PCA instead of PCA may lead to loss of discriminative information that could improve the accuracy of 2D-gMSM. This problem is addressed by E2D-PCA, where a radius of $r$ diagonals around the main diagonal of PCA is employed to construct the E2D-PCA scatter matrix. The parameter $r$ connects PCA and 2D-PCA, controlling the trade-offs between the basis vectors dimension and the recognition accuracy. Thus, E2D-PCA has a computational complexity ranging between the complexity of 2D-PCA and the complexity of conventional PCA.

## IV. Experimental Results

We conducted image set matching experiments on six datasets including the ALOI [8], RGB-D [9] for the object recognition task, Honda/UCSD [10], YouTube Celebrities (YTC) [11] and PubFig83 [12] for the face recognition. We compared the computational time and the classification rate of the proposed method with DCC [18], MMD [19], MDA [20] and CHISD [21] methods on the different datasets. For the test stage, we computed the processing time of classifying one image set with all training image sets. The performance we report is measured on a Unix-like PC equipped with a Core i7 2.2GHz quad core with 8 GB RAM under Matlab. For the experiments, we resized all the images to $40 \times 40$ pixels and, except from Color-gMSM, which can handle RGB data [15], all the other methods employed gray-scale images.

For the object recognition task, we employed ALOI dataset. ALOI is a large image database of general objects where the illumination angle, illumination color and the viewing angle, were systematically varied in order to produce about $110$ images for each object. In this experiment, we used the first $500$ object instances of the database. All images were segmented from the background and we classify an input set of images to one of the $500$ objects available in a $10-$fold cross validation scheme. We also employed RGB-D dataset, which consists of color and depth videos sequences of $300$ objects containing $51$ categories. The video sequences were taken from three different viewpoints. In our experiments, we subsample each sequence by taking every fifth frame, resulting in $41,877$ color and depth images. The objects in the dataset are already segmented from the background. We classify an input set of images in a $10-$fold cross validation scheme.

In Honda/UCSD dataset, we consider their first subset, which consists of $59$ videos of $20$ subjects. In each video, subject moves his face in an arbitrary sequence of 2-D and 3-D rotations while changing facial expression and speed, illumination conditions also vary significantly. Each video consists of about 300-500 frames and each subject has at least two videos. The face images were cropped and we classify an input set of images in a $10-$fold cross validation scheme by randomly selecting one sequence for each subject for training and using the rest for testing, as in [20]. For face recognition task, we also employed YTC dataset, which contains $1910$ video clips of $47$ celebrities, mostly actors and politicians, collected from YouTube under unconstrained conditions. Each video clip contains frames varying from $7$ to $400$. There are large variations of pose, illumination, and expression on face videos. In addition, the quality of face videos is very poor because most videos have high compression rate. This database is more challenging comparing to Honda/UCSD as the videos exhibit very large variations in face pose, illumination, expression, and other conditions. The face images were evaluated in a $10-$fold cross validation scheme.

PubFig83 dataset contains $8300$ cropped face images of $100 \times 100$ pixels, with $100$ images of each of $83$ subjects. There are large variations of pose, illumination, expression on face images because these images were captured in unconstrained environments from the Google images and FlickR. We also employ a $10-$fold cross validation scheme.

Table I lists the performances of 2D-gMSM (and variants) and gMSM in terms of the processing times (in seconds) and classification rate (%). We can observe that the classification time of 2D-gMSM (and most of its variants) are about $4$ times faster than the learning time and matching time of gMSM, revealing that the computational cost to obtain the subspaces from the covariance matrix employed by 2D-gMSM (and variants) is more efficient than the covariance matrix employed by gMSM. The scatter matrix of 2D-gMSM is formulated by sum of all scatter matrices of different column indices in the main diagonal of PCA. Hence, the processing times of 2D-gMSM and A2D-gMSM are very fast, comparing to gMSM because the number of features employed by these methods are less than the number of features employed by gMSM. However, replacing PCA by 2D-PCA and its variants, some discriminative features may be lost according to the compactness type of each method, decreasing its classification rate.

We can note that Color-gMSM achieved the highest recognition rate compared to the other methods. This is an expected result, since Color-gMSM can efficiently handle color information, which more discriminant features are available. Hence, the processing time of Color-gMSM is higher compared to gMSM and E2D-gMSM, due to its larger covariance matrix. E2D-gMSM has shown an interesting result, since its recognition rate is comparable to gMSM and its processing time is more efficient. E2D-gMSM has this behavior, due to the dynamic construction of its covariance matrix, which its size varies from the size of 2D-PCA and PCA. In our experiments, the parameter $r$ (the trade-offs between the subspace dimension and the classification rate) is set by experiments.

MDA and CHISD exhibited the higher classification rates (except from Color-gMSM) on YTC datasets. This result is due to the fact that MDA learns a linear discriminant function, maximizing the between-class manifolds separability, and achieving high classification rates on image face datasets. CHISD reduces the influence of outliers by applying robust methods to remove samples that do not fit the model. On the other hand, gMSM and the proposed methods do not apply robust techniques, only making use of the eigenvalues

TABLE I
PROCESSING TIME (SECONDS) OF DIFFERENT IMAGE SET CLASSIFICATION METHODS AND THE AVERAGE CLASSIFICATION RATES.

| Dataset | ALOI | | | RGB-D | | | Honda/UCSD | | | YTC | | | PubFig83 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | Train | Test | Class. Rate | Train | Test | Class. Rate | Train | Test | Class. Rate | Train | Test | Class. Rate | Train | Test | Class. Rate |
| DCC [18] | 93.9 | 2.3 | 90.1 ± 3.7 | 102.9 | 2.7 | 89.7 ± 2.4 | 58.1 | 1.6 | 92.8 ± 2.3 | 91.9 | 5.1 | 65.8 ± 4.5 | 24.5 | 1.6 | 45.5 ± 1.5 |
| MMD [19] | – | 3.9 | 85.8 ± 3.9 | – | 4.1 | 88.4 ± 2.6 | – | 4.3 | 92.6 ± 2.1 | – | 8.3 | 67.7 ± 3.8 | – | 2.9 | 46.3 ± 1.5 |
| MDA [20] | 117.1 | 4.1 | 90.2 ± 3.8 | 132.8 | 5.3 | 89.7 ± 2.5 | 83.9 | 2.9 | **94.5 ± 3.1** | 145.2 | 10.2 | **68.1 ± 4.3** | 67.1 | 2.5 | 48.6 ± 1.6 |
| CHISD [21] | – | 7.8 | 79.1 ± 4.2 | – | 10.4 | 85.2 ± 2.1 | – | 12.5 | 93.2 ± 2,1 | – | 27.2 | 67.4 ± 4.7 | – | 11.9 | 64.8 ± 2.1 |
| gMSM [5] | – | 3.5 | 91.2 ± 2.5 | – | 3.7 | 91.4 ± 1.9 | – | 5.6 | 94.1 ± 3.4 | – | 7.2 | 67.1 ± 4.8 | – | 5.3 | 64.7 ± 1.7 |
| 2D-gMSM | – | **0.9** | 86.6 ± 3.1 | – | **0.9** | 87.8 ± 2.1 | – | **0.9** | 89.7 ± 4.1 | – | **1.6** | 62.8 ± 5.1 | – | **0.9** | 60.4 ± 3.5 |
| A2D-gMSM | – | **0.9** | 86.5 ± 3.1 | – | **0.9** | 87.6 ± 2.3 | – | **0.9** | 88.9 ± 4.3 | – | **1.6** | 62.4 ± 4.3 | – | **0.9** | 60.2 ± 3.6 |
| E2D-gMSM | – | 1.6 | 91.2 ± 2.9 | – | 1.1 | 91.3 ± 2.2 | – | 1.3 | 93.9 ± 3.7 | – | 2.9 | 66.8 ± 4.9 | – | 1.3 | 64.5 ± 1.9 |
| Color-gMSM | – | 2.1 | **91.4 ± 2.7** | – | 1.9 | **91.7 ± 1.7** | – | 1.9 | 94.3 ± 2.1 | – | 4.0 | 67.3 ± 3.9 | – | 1.6 | **65.1 ± 1.5** |
| C2D-gMSM | – | 1.1 | 87.7 ± 3.4 | – | 1.1 | 88.1 ± 2.1 | – | 1.3 | 90.1 ± 3.9 | – | 2.1 | 63.3 ± 4.9 | – | 1.3 | 62.7 ± 2.9 |

(variance) to determine the importance of each eigenvector. In addition, gMSM and E2D-gMSM achieved reasonably competitive recognition rate on Honda/UCSD and Pub-Fig83 datasets. These methods are, therefore, robust enough to handle high variations on illumination conditions, camera angle and unconstrained backgrounds, inherent in such datasets.

It should be noted that many of the existing methods, as well as gMSM and the proposed methods, do not require training. Our proposed methods do not perform training and can adapt to newly added and previously unseen training data (e.g., when a new image set is included). However, one major limitation of our methods is that all the computation is done at run-time and comparatively more memory storage is required.

## V. CONCLUSIONS

A novel object recognition framework has been proposed for sets of images matching based on gMSM, 2D-PCA and variants. Most of the 2D-MSM variants are computational more efficient than the traditional method and the implementation are straightforward. The proposed E2D-gMSM and C2D-gMSM algorithms demonstrated an impressive processing time on all the evaluated databases and recognition rate equivalent to the traditional gMSM. Moreover, Color-gMSM exhibited the highest recognition rate on the same databases. We have shown that our proposed algorithms are theoretically and practically attractive. Our new approach speeds up the sets of images matching by creating a more compact representation from these sets due to 2D-PCA and variants inherent characteristics.

## REFERENCES


[1] J. Meltzer, M.-H. Yang, R. Gupta, and S. Soatto, "Multiple view feature descriptors from image sequences via kernel principal component analysis," in *Computer Vision-ECCV 2004*. Springer, 2004, pp. 215–227.

[2] K.-i. Maeda, "From the subspace methods to the mutual subspace method," in *Computer Vision*. Springer, 2010, pp. 135–156.

[3] E. Oja and E. Oja, *Subspace methods of pattern recognition*. Research Studies Press England, 1983, vol. 4.

[4] F. Chatelin, *Eigenvalues of Matrices: Revised Edition*. SIAM, 2012, vol. 71.

[5] T. Kobayashi, "Generalized mutual subspace based methods for image set classification," in *Computer Vision–ACCV 2012*. Springer, 2013, pp. 578–592.

[6] K. Fukui and A. Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 11, pp. 2164–2177, 2015.

[7] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, no. 1, pp. 131–137, 2004.

[8] J.-M. Geusebroek, G. J. Burghouts, and A. W. Smeulders, "The amsterdam library of object images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005.

[9] K. Lai, L. Bo, X. Ren, and D. Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1817–1824.

[10] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 1. IEEE, 2003, pp. I–313.

[11] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[12] N. Pinto, Z. Stone, T. Zickler, and D. Cox, "Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. IEEE, 2011, pp. 35–42.

[13] D. Zhang and Z.-H. Zhou, "(2d) 2pca: Two-directional two-dimensional pca for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1, pp. 224–231, 2005.

[14] M. Safayani, M. M. Shalmani, and M. Khademi, "Extended two-dimensional pca for efficient face representation and recognition," in *Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on*. IEEE, 2008, pp. 295–298.

[15] M. Thomas, S. Kumar, and C. Kambhamettu, "Face recognition using a color pca framework," in *Computer Vision Systems*. Springer, 2008, pp. 373–382.

[16] Ü. Turhal and A. Duysak, "Cross grouping strategy based 2dpca method for face recognition," *Applied Soft Computing*, vol. 29, pp. 270–279, 2015.

[17] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, 2006.

[18] T.-K. Kim, J. Kittler, and R. Cipolla, "Learning discriminative canonical correlations for object recognition with image sets," in *Computer Vision–ECCV 2006*. Springer, 2006, pp. 251–262.

[19] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[20] R. Wang and X. Chen, "Manifold discriminant analysis," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE, 2009, pp. 429–436.

[21] H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2567–2573.

[22] H. Murase and S. K. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International journal of computer vision*, vol. 14, no. 1, pp. 5–24, 1995.