ROTATION INVARIANT SIMULTANEOUS CLUSTERING AND DICTIONARY LEARNING

Yi-Chen Chen[1], *Challa S. Sastry*[2], *Vishal M. Patel*[1], *P. Jonathon Phillips*[3], *and Rama Chellappa*[1]

[1]Department of Electrical and Computer Engineering
Center for Automation Research, University of Maryland, College Park, MD
[2]Department of Mathematics, Indian Institute of Technology Hyderabad, Hyderabad, India
[3]National Institute of Standards and Technology, Gaithersburg, MD

{chenyc08,pvishalm,rama}@umiacs.umd.edu csastry@iith.ac.in jonathon.phillips@nist.gov

ABSTRACT

In this paper, we present an approach that simultaneously clusters database members and learns dictionaries from the clusters. The method learns dictionaries in the Radon transform domain, while clustering in the image domain. The main feature of the proposed approach is that it provides rotation invariant clustering which is useful in Content Based Image Retrieval (CBIR). We demonstrate through experimental results that the proposed rotation invariant clustering provides better retrieval performance than the standard Gabor-based method that has similar objectives.

Index Terms— Radon transform, rotation invariance, clustering, dictionary learning, CBIR.

1. INTRODUCTION

In recent years, sparse representation has emerged as a powerful tool for efficiently processing data in non-traditional ways. This is mainly due to the fact that signals and images of interest tend to enjoy the property of being sparse in some dictionary [1]. These dictionaries are often learned directly from the training data. It has been observed that learning dictionaries directly from examples usually leads to improved results in many practical image processing applications such as restoration and classification [2]. One of the simplest algorithms for learning dictionaries is the K-SVD algorithm [2]. Given a set of images $\{\mathbf{x}_i\}_{i=1}^n$, K-SVD seeks a dictionary $\mathbf{D}$ that provides the sparsest representation for each image by solving the following optimization problem

$$(\hat{\mathbf{D}}, \hat{\mathbf{\Gamma}}) = \arg\min_{\mathbf{D},\mathbf{\Gamma}} \|\mathbf{X} - \mathbf{D}\mathbf{\Gamma}\|_F^2 \text{ subject to } \forall i \ \|\boldsymbol{\gamma}_i\|_0 \leq T_0, \quad (1)$$

where $\boldsymbol{\gamma}_i$ represents the i^{th} column of $\mathbf{\Gamma}$, $\mathbf{X}$ is the matrix whose columns are $\mathbf{x_i}$ and T_0 is the sparsity parameter. Here, $\|\mathbf{A}\|_F$ denotes the Frobenius norm defined as $\|\mathbf{A}\|_F = \sqrt{\sum_{ij} \mathbf{A}_{ij}^2}$. The K-SVD algorithm alternates between sparse-coding and dictionary update steps. In the sparse-coding step, $\mathbf{D}$ is fixed and the representation vectors $\boldsymbol{\gamma}_i$s are found for each example $\mathbf{x}_i$. Then, the dictionary is updated atom-by-atom in an efficient way.

While these dictionaries are often trained to obtain good reconstruction, training dictionaries with a specific discriminative criterion has also been considered. For instance, linear discriminant analysis (LDA) based basis selection and feature extraction algorithm for classification using wavelet packets was proposed by Etemand and Chellappa [3] in the late nineties. Recently, similar algorithms for simultaneous sparse signal representation and discrimination have also been proposed [4], [5], [6] ,[7], [8], and [9]. Additional techniques may be found within these references.

This work was partially supported by an ONR grant N00014-12-1-0124.

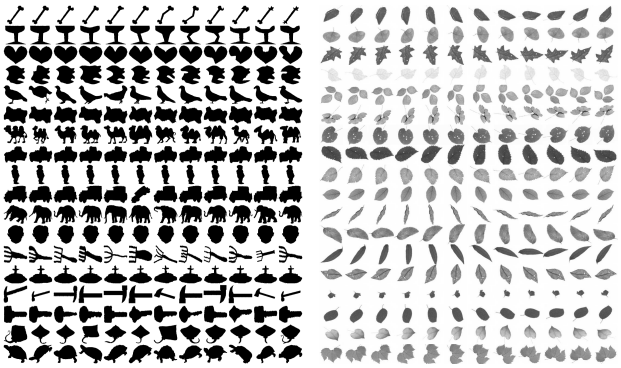


Fig. 1. Left: Kimia's database. Right: Smithsonian isolated leaf database with rotated images.

Dictionary learning techniques for unsupervised clustering have also gained some traction in recent years. In [10], a method for simultaneously learning a set of dictionaries that optimally represent each cluster is proposed. To improve the accuracy of sparse coding, this approach was later extended by adding a block incoherence term in the optimization problem [11]. Some of the other sparsity motivated subspace clustering methods include [12], [13].

Rotation invariance is an important property in many applications such as image classification and retrieval where one wants to classify or retrieve images having same content but different orientation. For instance, in content based image retrieval (CBIR), images are retrieved from a database using features that best describe the orientation of objects in the query image. Due to the ability of Gabor filters to capture directional information, they are often used to extract features for retrieval. However, the chosen directions of Gabor filters may not correspond to the orientation of the content in the query image. Hence, a feature extraction method that is independent of orientation in the image is desirable. Fig. 1 shows some sample images from the Kimia's object dataset and Smithsonian leaf dataset, where the presence of orientation is clearly seen in the images.

In this paper, we present a rotation invariant clustering method, extending the dictionary learning and sparse representation framework for clustering of data. Given a database of images $\{\mathbf{x}_j\}_{j=1}^N$ and the number of clusters K, we learn K dictionaries for representing the data. The dictionaries are learnt in the Radon transform

domain which ensures that the clustering is rotation independent. We demonstrate the effectiveness of our approach in retrieval experiments, where significant improvements are shown.

The organization of the paper is as follows. Our rotation invariant clustering framework is detailed in Section 2. We demonstrate experimental results in Section 3 and Section 4 concludes the paper with a brief summary and discussion.

2. SIMULTANEOUS CLUSTERING AND DICTIONARY LEARNING

In this section, we present our rotation invariant clustering method. We first discuss how Radon transform is used to detect rotation present in an image.

2.1. Estimating the rotation present in images

The Radon transform of a sufficiently regular continuous domain two variable function x is defined as

$$R_\theta x(t) = \int_{-\infty}^{\infty} x(t\cos\theta - s\sin\theta, t\sin\theta + s\cos\theta)ds, \qquad (2)$$

where $(t,\theta) \in (-\infty,\infty)\times[0,\pi)$. If $\tilde{x}$ is the rotated copy of x by an angle $\hat{\theta}$, then a simple proof shows that their Radon transforms are related as

$$R_\theta \tilde{x}(t) = R_{\theta+\hat{\theta}}x(t), \ \ \forall t,\theta. \qquad (3)$$

The variance σ_θ of the Radon transform is found [14] to be useful in estimating the presence of angle in images. Therefore, given the image $\tilde{x}$, one may estimate $\hat{\theta}$ from the following formula

$$\hat{\theta} = \arg\min_\theta \left(\frac{d^2\sigma_\theta}{d\theta^2}\right). \qquad (4)$$

An application of properties of the Fourier transforms implies that $\hat{\theta}$ can be estimated from the following simple formula

$$\hat{\theta} = \arg\min_\theta \left(\frac{d^2}{d\theta^2}\int_{-\infty}^{\infty} X^2(s\cos\theta, s\sin\theta)ds\right), \qquad (5)$$

where X is the Fourier transform of x. In all practical situations, for an image of size $m\times n$, the Radon transform is represented as a matrix $[\mathbf{R}_{\theta_l} f(t_m)]$, called *sinogram*. Usually, one takes $\theta_l = \frac{(l+0.5)\pi}{n}, l = 0,1,\ldots,n-1$ and $t_m = \delta r(k-\frac{m}{2}), k = 0,1,\ldots,m-1$, where δr is the radial sampling, which is chosen to avoid aliasing error in the reconstruction.

The second image shown on the first row of Fig. 2 [1] is a rotated copy of the first image by 30°. The plots on the second row are the second derivatives of variances of Radon transform of them. It may be noted that the difference between the points of global minima of both curves is 30, coinciding with the rotation present in the second image. Consequently, the estimate presented in (5) is useful in estimating the presence of rotation in the images.

2.2. Proposed algorithm

Let $\{\mathbf{x}_j\}_{j=1}^N$ be the database of images and K be the number of clusters. Define $\mathbf{D} = [\mathbf{D}_1,\ldots,\mathbf{D}_K]$ as the concatenation of dictionaries corresponding to K clusters. Let $\mathbf{C}_i$ be the matrix containing images as columns corresponding to cluster $\mathbb{C}_i$. Equipped with the above notation and motivated by dictionary learning methods [2], we realize our objective in two steps

[1] http://www.flowersofpictures.com/wp-content/uploads/2011/07/Lily-Flowers.jpg

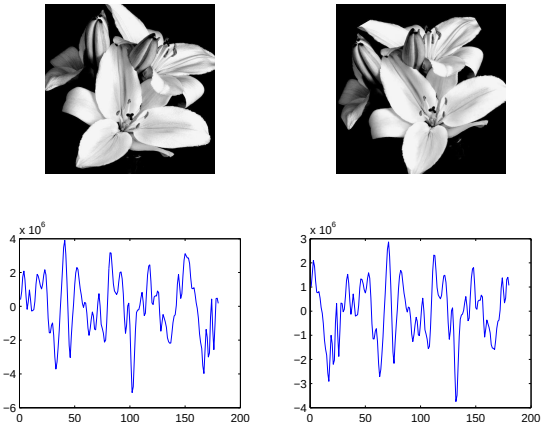


Fig. 2. For the rotated images present on the first row, the plots on the second row are their $\frac{d^2\sigma_\theta}{d\theta^2}$. The second row plots indicate that the difference between the points of global minimum of both curves preserves the rotation present in the second image.

- *Cluster assignment:* We start with arbitrary dictionary $\mathbf{D} = [\mathbf{D}_1,\ldots,\mathbf{D}_K]$. Given an image $\mathbf{x}_j$ and its estimated orientation $\hat{\theta}_j$ which is calculated from the discretized version of (5), we use $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ to denote the Radon transform matrix of the rotated version of the image $\mathbf{x}_j$ by $-\hat{\theta}_j$. In other words, $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ is an aligned Radon transform matrix of $\mathbf{x}_j$. It is obtained by left shifting columns of the original Radon transform matrix of $\mathbf{x}_j$ by $\hat{\theta}_j$. Our approach considers obtaining the sparsest representation of $\mathbf{R}_{\hat{\theta}_j}\mathbf{x}_j$ in an appropriate dictionary $\mathbf{D}_{\hat{i}}$ from

$$\begin{aligned}\boldsymbol{\alpha}^j &= \arg\min_{\boldsymbol{\omega}} \|\mathbf{R}_{\hat{\theta}_\mathbf{j}}\mathbf{x_j} - \mathbf{D}\boldsymbol{\omega}\|_2^2 \ \text{ s. t. } \ \|\boldsymbol{\omega}\|_0 \le T_0,\\ \hat{i} &= \arg\min_i \|\mathbf{R}_{\hat{\theta}_\mathbf{j}}\mathbf{x_j} - \mathbf{D}\delta_i(\boldsymbol{\alpha}^j)\|_2^2 \ \ j = 1,\cdots,N,\end{aligned} \qquad (6)$$

  then $\mathbf{x}_j$ is set to belong to $\mathbb{C}_{\hat{i}}$. In (6), δ_i is a characteristic function that selects the coefficients associated with the i^{th} dictionary.

- *Dictionary update:* Having obtained the initial clusters $\mathbb{C}_1,\mathbb{C}_2,\ldots,\mathbb{C}_K$, we update the dictionaries $\mathbf{D}_i$ using the K-SVD approach described in the previous section. The new dictionaries are obtained by solving the following optimization problem

$$(\hat{\mathbf{D}}_\mathbf{i},\hat{\mathbf{\Gamma}}_\mathbf{i}) = \arg\min_{\mathbf{D}_i,\mathbf{\Gamma_i}} \|\mathbf{C_i} - \mathbf{D}_i\mathbf{\Gamma_i}\|_F^2 \text{ s.t } \forall i \ \|\boldsymbol{\gamma}_i\|_0 \le T_0,$$

  satisfying $\mathbf{C}_i = \hat{\mathbf{D}}_i\hat{\mathbf{\Gamma}}_i, \quad i = 1,2,\ldots,K.$

We repeat the cluster assignment and dictionary update steps till there is no significant change in $\mathbb{C}_i$. It may be noted here that the dictionaries are learnt in the Radon domain, while clustering is done in the image domain. The clustering methodology presented above may be summarized in terms of the following optimization problem

$$\min_{\mathbf{D}_i,\mathbb{C}_i}\sum_{i=1}^K\sum_{\mathbf{x}\in\mathbb{C}_i}\min_\theta\left\{\|\mathbf{R}_\theta\mathbf{x} - \mathbf{D}\delta_i(\alpha)\|_2^2 + \gamma\|\alpha\|_1 + \mu\left|\frac{d^2\sigma_\theta}{d\theta^2}\right|\right\}, \qquad (7)$$

where $\gamma,\mu > 0$. Here, σ_θ is the variance of the first column of $\mathbf{R}_\theta\mathbf{x}$. The last term in (7) estimates the presence of rotation in images.

We refer to our rotation invariant clustering and dictionary learning method as RICD.

2.3. Application to CBIR

Once the dictionaries have been learnt for each class in Radon domain, given a query image $\mathbf{x}_q$, we estimate $\hat{\theta}_q$ and then project $\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q$ onto the span of the atoms in each $\mathbf{D}_i$ using the orthogonal projector

$$\mathbf{Proj}_{\mathbf{D}_i} = \mathbf{D}_i(\mathbf{D}_i^T\mathbf{D}_i)^{-1}\mathbf{D}_i^T. \quad (8)$$

The approximation and residual vectors can then be calculated as

$$\mathbf{R}^i_{\hat{\theta}_q}\mathbf{x}_q = \mathbf{Proj}_{\mathbf{D}_i}(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q), \quad (9)$$

and

$$\mathbf{r}^i(\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q) = \mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q - \mathbf{R}^i_{\hat{\theta}_q}\mathbf{x}_q = (\mathbf{I} - \mathbf{Proj}_{\mathbf{D}_i})\mathbf{R}_{\hat{\theta}_q}\mathbf{x}_q, \quad (10)$$

respectively, where $\mathbf{I}$ is the identity matrix. Since the dictionary learning step in our algorithm finds the dictionary, $\mathbf{D}_i$, that leads to the best representation for each member of $\mathbb{C}_i$ in Radon domain, we suspect $\|\mathbf{R}_{\hat{\theta}_q,i}\mathbf{x}_q\|_2$ to be small if $\mathbf{x}_q$ were to be relevant to the i^{th} cluster and large for the other clusters. Based on this, if

$$d = \arg\min_{1\le i\le K} \|\mathbf{r}^i(\mathbf{R}_{\hat{\theta}_q,i}\mathbf{x}_q)\|_2,$$

we search for the relevance of $\mathbf{x}_q$ in $\mathbb{C}_d$.

3. EXPERIMENTAL RESULTS

To demonstrate the effectiveness of our method, we present experimental results on various datasets [15], [16],[17]. We evaluate the performance of our method using precision-recall curves and average retrieval performance [18]. Recall and precision are defined as

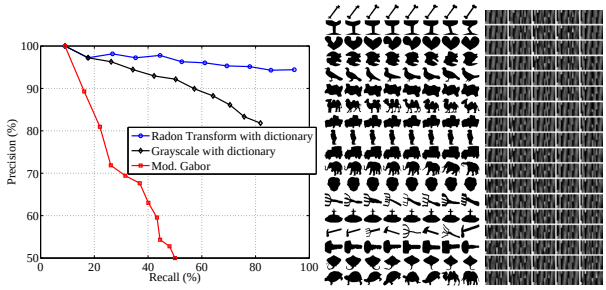
$$\text{Recall} = \frac{\text{Number of relavant images retrieved}}{\text{Total number of relevant images}}, \quad (11)$$

$$\text{Precision} = \frac{\text{Number of relavant images retrieved}}{\text{Total number of images retrieved}}. \quad (12)$$

Recall is the portion of total relevant images retrieved whereas precision indicates the capability to retrieve relevant only images. An ideal retrieval should give precision rate which always equals 1 for any recall rate. Given a certain number of retrieved images, the average retrieval performance is defined as the average number of relevant retrieved images over all query images of a particular class. We compare the performance of our method with that of modified Gabor-based approach [18] and [10]. Note that [18] uses features that are rotation invariant and the method presented in [10] uses a discriminative dictionary learning approach to clustering. We refer to the method presented in [10] as dictionary-based clustering (DC). In all the experiments, the dictionaries are initialized by randomly partitioning the Radon transformed data into K subsets.

3.1. Kimia database

This database consists of 216 images of 18 shapes. Each shape has 12 different images. Fig. 1(a) shows sample images from this database. The images were resized to 100×80. The precision-recall curves are shown in Fig. 3(a) for 18 query images belonging to different classes. Our Radon transform-based approach is able to maintain precision rate above 94% even when the recall rate reaches 90%. The modified Gabor-based method's precision goes down to 50% at approximately half recall rate. It is clearly seen from the figure that the performance of our method is better than the other methods. Fig. 3(b) shows 18 clusters obtained using our method. A few atoms from the learned dictionaries from each shape are shown in Fig. 3(c).



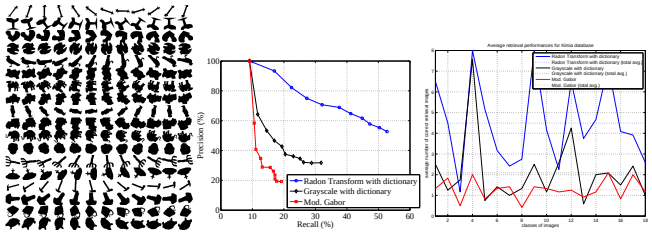
(a) (b) (c)

Fig. 3. Results on Kimia's database. (a) Precision-recall graphs. (b) Clustering results. (c) Associated learned dictionaries.

The original Kimia's dataset contains various shapes with small rotation. In order to create a more challenging dataset, we selected one representative image from each of the 18 shapes. For each selected shape, we created 11 in-plane rotated images with the following angles:

$$18^\circ, 36^\circ, 54^\circ, 72^\circ, 90^\circ, 108^\circ, 126^\circ, 144^\circ, 162^\circ, 180^\circ, 198^\circ.$$

The resulting dataset (shown in Fig. 4(a)) is more challenging than the original Kimia's dataset as it contains in-plane rotated images with various angles. The precision-recall curves are shown in Fig. 4(b) for 18 query images belonging to different shapes. Fig. 4(c) shows the total average retrieval performance over all shapes. On average our method obtained 4.4907 out of 8 retrieved images per shape. Whereas the DC method and Gabor-based method obtained 2.0926 and 1.2778, respectively. As can be seen from the figures, our method performed the best.



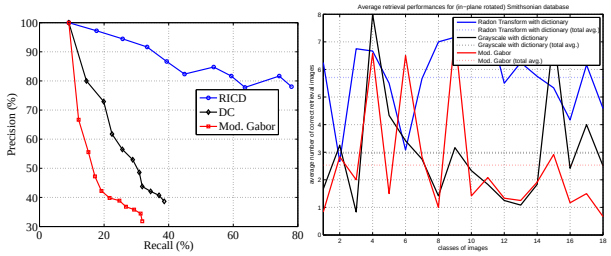
(a) (b) (c)

Fig. 4. Results on in-plane rotated Kimia's database. (a) Database. (b) Precision-recall curves. (c) Average retrieval performance.

3.2. Smithsonian isolated leaf database

The original Smithsonian isolated leaf database consists of 93 different leaves [16]. We selected one representative image from each of the last 18 leaves. As before, for each representative image, we created 11 in-plane rotated images with the same angles as considered in the previous experiment. Fig. 1(b) shows the resulting database containing 18 different leaves with rotated images. The images were resized to 100×80. This database is more challenging than the

Kimia database as more shape similarities can be easily found among different leaves (for instance, 5th and 6th rows; 9th and 10th rows; 11th and 13th rows in Fig. 1(b).) Fig. 5(a) shows the performance of different methods on this database. Fig. 5(b) shows the total average retrieval performance over all shapes. On average our method obtained 5.7083 out of 8 retrieved images per shape. Whereas the DC method and Gabor-based method obtained 2.9630 and 2.5324, respectively. Even on this difficult dataset, our method performs better than the other methods.



(a) (b)

Fig. 5. Experiment with in-plane rotated Smithsonian isolated leaf database. (a) Precision-recall graphs. (b) Average retrieval performance.

3.3. PIE database

Even though, our method can provide rotation invariant clustering, it can be generalized to handle other variations present in the data by using appropriate features. To illustrate this, we consider the frontal face images of the CMU PIE dataset. Our objective is to show the effectiveness of our approach in retrieving face images in the presence of various illumination conditions. We use the principle component analysis (PCA) features for this experiment. This database consists of 68 subjects, each of which contains 21 images under various illumination conditions. Eighteen subjects from this dataset are used for this experiment. All images are resized to 48×40. Fig. 6 shows the retrieval results of our method. The top row shows 18 query images from 18 different subjects. For each query image, the corresponding column shows the first six retrieved images. The false retrievals are marked with red boxes. As can be seen from this figure, there is no false retrieval from the first rank up to the third rank matches. In this experiment, the average retrieval performance is $(1 - \frac{8}{108}) \times 6 = 5.5556$ out of 6 images.

Fig. 6. Retrieval results on the PIE database.

4. CONCLUSION

In this paper, we have proposed a rotation invariant clustering algorithm suitable for such applications as content based image retrieval. With a view to achieving rotation invariance in clustering, the method learns dictionaries in the Radon transform domain though, the simultaneous clustering is done in the image domain. We demonstrated the effectiveness of the proposed method for CBIR applications.

5. REFERENCES

[1] P. J. Phillips, "Matching pursuit filters applied to face identification," *IEEE Trans. Image Process.*, vol. 7, no. 8, pp. 150–164, 1998.

[2] M. Aharon, M. Elad, and A. M. Bruckstein, "The k-svd: an algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.

[3] K. Etemand and R. Chellappa, "Separability-based multiscale basis selection and feature extraction for signal and image classification," *IEEE Transactions on Image Processing*, vol. 7, no. 10, pp. 1453–1465, Oct. 1998.

[4] F. Rodriguez and G. Sapiro, "Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries," *Tech. Report, University of Minnesota*, Dec. 2007.

[5] K. Huang and S. Aviyente, "Sparse representation for signal classification," *NIPS*, vol. 19, pp. 609–616, 2007.

[6] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 2691–2698, 2010.

[7] M. Ranzato, F. Haung, Y. Boureau, and Y. LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[8] J. Mairal, F. Bach, J. Pnce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," *Proc. of the Conference on Computer Vision and Pattern Recognition*, Anchorage, AL, June 2008.

[9] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," *Advances in Neural Information Processing Systems*, Vancouver, B.C., Canada, Dec. 2008.

[10] P. Sprechmann and G. Sapiro, "Dictionary learning and sparse coding for unsupervised clustering," in *ICASSP*, march 2010, pp. 2042 –2045.

[11] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *CVPR*, june 2010, pp. 3501 –3508.

[12] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *CVPR*, june 2009, pp. 2790 –2797.

[13] S. Rao, R. Tron, R. Vidal, and Y. Ma, "Motion segmentation via robust subspace separation in the presence of outlying, incomplete, or corrupted trajectories," in *CVPR*, june 2008, pp. 1 –8.

[14] K. Jafari-Khouzani and H. Soltanian-Zadeh, "Radon transform orientation estimation for rotation invariant texture analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 1004–1008, 2005.

[15] T. B. Sebastian, P. N. Klein, and B. B. Kimia, "Recognition of shapes by editing their shock graphs," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 26, no. 5, pp. 550–571, 2004.

[16] H. Ling and D. W. Jacobs, "Shape classification using the inner-distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, pp. 286–299, 2007.

[17] S. Baker, "The cmu pose, illumination, and expression (PIE) database (http://www.ri.cmu.edu/projects/project_418.html)."

[18] C. S. Sastry, M. Ravindranath, A. K. Pujari, and B. Deekshatulu, "A modified Gabor function for content based image retrieval," *Pattern Recognition Letters*, pp. 293–300, 2007.