21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Robust Motion Segmentation via Refined Sparse Subspace Clustering

Hao Ji[1], Fei Su[1,2]
[1] *School of Information and Communication Engineering*
[2] *Beijing Key Laboratory of Network System and Network Culture*
Beijing University of Posts and Telecommunications, Beijing, China

Abstract

In this paper, a new refined sparse subspace clustering (RSSC) method is proposed for robust motion segmentation. Given a set of trajectories of tracked feature points from multiple moving object, RSSC aims at seeking a sparse representation (SR) for each trajectory with respect to a recovered low-rank dictionary. The segmentation of motion is obtained by applying spectral clustering to the affinity matrix built by this SR. Compared to the conventional sparse subspace clustering (SSC) algorithm, our RSSC integrates sparse representation and low-rank subspace structures recovery into a unified framework. Furthermore, SR is obtained from the recovered dictionary instead of the initial given dictionary built by contaminated data, making RSSC more robust to data noise. Experiments on toydata and real video sequences (Hopkins 155 database) show the superiority of our approach over several current state of the art methods.

1. Introduction

Motion segmentation is a hot research topic in computer vision due to its significant role in video analysis and understanding. While traditional research focuses on the static scenes, there has been a growing interest in the analysis of dynamic scenes. Such scenes may contain multiple motions, including objects motion and background motion due to camera moving. As illustrated in Fig.1, the basic task of motion segmentation is to segment a set of trajectories of tracked points into different groups corresponding to respective motions without any prior or label information.

Under the affine camera model, all the trajectories associated with a single rigid object live in a 3 dimensional affine subspace. Given tracked points trajectories $\left\{x_{fp} \in \mathbb{R}^2\right\}_{p=1,\dots,P}^{f=1,\dots,F}$ in F 2-D frames of P 3-D points

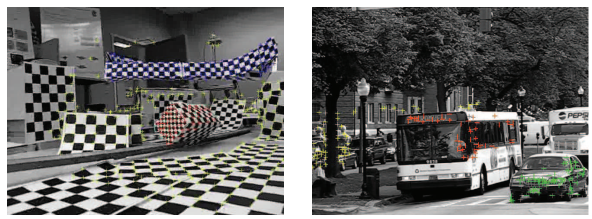

Figure 1: Frames from two example video sequences involving three motions. Left: two moving checkerboards plus background. Right: two moving automobiles plus background.

$\left\{\mathbf{X}_p \in \mathbb{R}^3\right\}_{p=1,\dots P}$ on a rigidly moving object. The relationship between the tracked points and the corresponding 3-D coordinates is

$$\begin{bmatrix} x_{11}\cdots x_{1P} \\ \vdots \\ x_{F1}\cdots x_{FP} \end{bmatrix}_{2F\times P} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_F \end{bmatrix} \begin{bmatrix} \mathbf{X}_1\cdots \mathbf{X}_P \\ 1\cdots 1 \end{bmatrix} \quad (1)$$

where x_{fp} is the 2-D coordinates of point p at frame f, $\mathbf{A}_f \in \mathbb{R}^{2\times 4}$ is the affine motion matrix at frame f and $\mathbf{X}_p$ is the 3-D coordinates of point p. As the rank of right hand side of Eq.1 is at most 4, all columns of the matrix of the left hand side must lie in a 4-dimensional subspace of $\mathbb{R}^{2F}$. Assume now we are given P trajectories from n rigidly moving objects (the rank is bounded by $4n \ll 2F$), the motion segmentation problem reduces to a subspace clustering process [1].

Existing works on motion segmentation can be roughly divided into four categories: iterative, factorization, algebraic and spectral clustering. Iterative approaches, such as K-subspaces and RANSAC [2], iteratively find a min-max estimation. Factorization-based methods [3] find a data segmentation from the similarity matrix built by factorization. Algebraic methods, such as Generalized Principal Component Analysis (GPCA) [4], fit data with a polynomial model whose gradient gives the normal vector to the subspaces. In order to achieve robustness to noise, Sparse Subspace Clustering

978-4-9906441-1-6 ©2012 IAPR

(SSC) [5], as a spectral clustering way, uses the sparsest representation from given dictionary to construct the 'block-sparse' affinity matrix. However, utilizing the initial data matrix as dictionary directly may cause inaccuracy due to corrupted data.

In this paper, we present a Refined Sparse Subspace Clustering (RSSC) method to address the contaminated dictionary problem. Compared to initial SSC, RSSC aims at removing noise and outliers from data and performing robust subspace clustering simultaneously. Since the low-rank nature of trajectories matrix (at most $4n$), recovery of dictionary is performed by adding a nuclear norm regularization to the objective function. Meanwhile, the 'block-sparse' attribute of affinity matrix is also preserved by using sparse representation driven by ℓ^1-norm. We also propose an efficient algorithm based on inexact Augmented Lagrange Multiplier (IALM) method [6] to solve the RSSC optimization model.

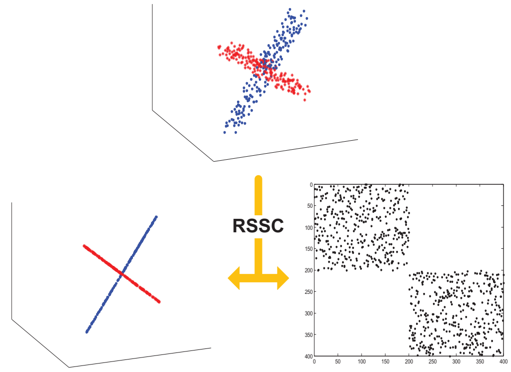

Figure 2: The process of RSSC in a mixture of two 1D subspaces. Top fig shows the data sampled from subspaces with noise, bottom figs shows the recovered data and the corresponding 'block-sparse' representation by RSSC.

2. Robust Motion Segmentation via RSSC

In this section, we will first briefly review the initial SSC method for subspace clustering. Then our Refined Sparse Subspace Clustering (RSSC) approach is proposed to improve the robustness to noises. Finally a efficient IALM method is presented to solve the optimization problem.

2.1 Sparse Subspace Clustering

SSC [5] uses the sparsest representation with respect to a given dictionary to construct the 'block-sparse' affinity matrix. Such a representation $\mathbf{R}$ can be obtained by solving the following program

$$\begin{aligned} &\min_{\mathbf{R}} \tfrac{1}{2}\|\mathbf{D}-\mathbf{D}\mathbf{R}\|_F^2+\lambda\|\mathbf{R}\|_1 \\ &\text{s.t. } \mathbf{R}_{ii}=0 \quad i=1,2\cdots,P \end{aligned} \tag{2}$$

where $\|.\|_F$ is F-norm, $\|.\|_1$ is ℓ^1-norm, $\mathbf{D}\in\mathbb{R}^{K\times P}$ is the data matrix, $\mathbf{R}\in\mathbb{R}^{P\times P}$ is the sparse representation and is used as the affinity matrix, $\lambda>0$ is a parameter. However, utilizing the data matrix as the given dictionary directly may cause inaccuracy because data points drawn from a union of subspaces are always contaminated by noise.

2.2 Refined Sparse Subspace Clustering

As illustrated in Fig.2, Refined Sparse Subspace Clustering (RSSC) aims at simultaneously recovering the corrupted dictionary and obtaining the sparse representation. Recovered data are used instead of the noisy data for SR. Due to the low-rank nature of data matrix (bounded by $4n$, n is the number of subspaces or motions), we assume that $\mathbf{D}=\mathbf{X}+\mathbf{E}$, where $\mathbf{D}$ is the contaminated data matrix, $\mathbf{X}$ is the recovered low-rank data matrix and $\mathbf{E}$ denotes noises or outliers. Therefore, rather than directly obtaining sparse representation from noisy dictionary $\mathbf{D}$, we get the 'block-sparse' affinity matrix based on the clean dictionary $\mathbf{X}$. Following these perspectives, we get the following optimization model for robust motion segmentation problem:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{R},\mathbf{E}} &\|\mathbf{E}\|_\ell+\mu_1 rank(\mathbf{X})+\mu_2\|\mathbf{R}\|_0 \\ \text{s.t. } &\mathbf{D}=\mathbf{X}+\mathbf{E} \\ &\mathbf{X}=\mathbf{X}\mathbf{R} \\ &\mathbf{R}_{ii}=0 \quad i=1,2\cdots,P \end{aligned} \tag{3}$$

where $\|.\|_\ell$ is some kind of regularization strategy, such as F-norm for noise or reconstruction error, ℓ^1-norm for outliers. In the following paper, we only consider the F-norm case of regularization strategy. The above optimization problem is non-convex and NP-hard. Fortunately, as suggested by compressive sensing (CS) theory [7] and matrix completion method [8], the following convex optimization provides a good surrogate for Eq.3:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{R}} &\tfrac{1}{2}\|\mathbf{D}-\mathbf{X}\|_F^2+\mu_1\|\mathbf{X}\|_*+\mu_2\|\mathbf{R}\|_1 \\ \text{s.t. } &\mathbf{X}=\mathbf{X}\mathbf{R} \\ &\mathbf{R}_{ii}=0 \quad i=1,2\cdots,P \end{aligned} \tag{4}$$

where $\|.\|_*$ denotes nuclear norm (sum of singular values) of a matrix.

Algorithm 1 Solving Problem (4) by Inexact ALM

Input: contaminated data matrix $\mathbf{D}$

Initialize: $\mathbf{X} = \mathbf{Q} = \mathbf{D}$, $\mathbf{R} = \mathbf{L} = 0$, $\mathbf{Y}_1 = 0$, $\mathbf{Y}_2 = 0$, $\mathbf{Y}_3 = 0$, $\beta = 10^{-6}$, $max_\beta = 10^{10}$, $\rho = 1.1$, $\epsilon = 10^{-6}$.

while not converged **do**

1: Update $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{L}$ and $\mathbf{X}$ respectively

$\mathbf{Q} = \arg\min \frac{1}{2}\|\mathbf{Q} - (\mathbf{X} + \mathbf{Y}_2/\beta)\|_F^2 + \frac{\mu_1}{\beta}\|\mathbf{Q}\|_*$

$\mathbf{R} = \frac{1}{\beta}(\mathbf{X}^T\mathbf{X}+\mathbf{I})^{-1}\left[\mathbf{X}^T\mathbf{Y}_1 - \mathbf{Y}_3 + \beta(\mathbf{X}^T\mathbf{X} + \mathbf{L})\right]$

$\mathbf{L} = \arg\min \frac{1}{2}\|\mathbf{L} - (\mathbf{R} + \mathbf{Y}_3/\beta)\|_F^2 + \frac{\mu_2}{\beta}\|\mathbf{L}\|_1$
s.t. $\mathbf{L}_{ii} = 0 \quad i = 1, 2 \cdots, P$

$\mathbf{X} = \left[\mathbf{D} - \mathbf{Y}_1(\mathbf{I} - \mathbf{R})^T - \mathbf{Y}_2 + \beta\mathbf{Q}\right]\{\mathbf{I} + \beta[(\mathbf{I} - \mathbf{R})(\mathbf{I} - \mathbf{R})^T + \mathbf{I}]\}^{-1}$

2: Update the multipliers

$\mathbf{Y}_1 = \mathbf{Y}_1 + \beta(\mathbf{X} - \mathbf{XR})$
$\mathbf{Y}_2 = \mathbf{Y}_2 + \beta(\mathbf{X} - \mathbf{Q})$
$\mathbf{Y}_3 = \mathbf{Y}_3 + \beta(\mathbf{R} - \mathbf{L})$

3: Update the parameter $\beta = min(\rho\beta, max_\beta)$

4: Check the convergence conditions

$\|\mathbf{X} - \mathbf{XR}\|_\infty < \epsilon$, $\|\mathbf{X} - \mathbf{Q}\|_\infty < \epsilon$, $\|\mathbf{R} - \mathbf{L}\|_\infty < \epsilon$

end while

2.3 Solving the RSSC by IALM

In this section, we propose an efficient algorithm based on inexact Augmented Lagrange Multiplier (IALM) method [6] to solve the above RSSC optimization model. We first convert Eq.4 to the following equivalent problem by introducing two auxiliary variables $\mathbf{Q}$ and $\mathbf{L}$:

$$\begin{aligned} \min_{\mathbf{X},\mathbf{R},\mathbf{Q},\mathbf{L}} \ & \frac{1}{2}\|\mathbf{D} - \mathbf{X}\|_F^2 + \mu_1\|\mathbf{Q}\|_* + \mu_2\|\mathbf{L}\|_1 \\ \text{s.t.} \ & \mathbf{X} = \mathbf{XR} \\ & \mathbf{X} = \mathbf{Q} \\ & \mathbf{R} = \mathbf{L} \\ & \mathbf{L}_{ii} = 0 \quad i = 1, 2 \cdots, P \end{aligned} \tag{5}$$

which can be solved by minimizing the following augmented Lagrange function:

$$\begin{aligned} & \mathcal{J}(\mathbf{X}, \mathbf{R}, \mathbf{Q}, \mathbf{L}, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3) \\ & = \frac{1}{2}\|\mathbf{D} - \mathbf{X}\|_F^2 + \mu_1\|\mathbf{Q}\|_* + \mu_2\|\mathbf{L}\|_1 \\ & + \langle \mathbf{Y}_1, \mathbf{X} - \mathbf{XR}\rangle + \langle \mathbf{Y}_2, \mathbf{X} - \mathbf{Q}\rangle + \langle \mathbf{Y}_3, \mathbf{R} - \mathbf{L}\rangle \\ & + \frac{\beta}{2}(\|\mathbf{X} - \mathbf{XR}\|_F^2 + \|\mathbf{X} - \mathbf{Q}\|_F^2 + \|\mathbf{R} - \mathbf{L}\|_F^2), \end{aligned} \tag{6}$$

where $\mathbf{Y}_i, i = 1, 2, 3$ are Lagrangian multipliers and $\beta > 0$ is penalty parameters for the constrains. IALM is derived by successively minimizing the augmented Lagrangian function $\mathcal{J}$ with respect to $\mathbf{X}$, $\mathbf{R}$, $\mathbf{Q}$ and $\mathbf{L}$, which is outlined in Algorithm 1. The affinity matrix can be obtained by $\mathbf{R} + \mathbf{R}^T$ after solving problem 4. Then spectral clustering algorithm is used on the affinity matrix to produce the final segmentation results.

3. Experiments

3.1 Toy Data

In this experiment, we demonstrate the robustness of our RSSC method to contaminated data and the recovery ability by low-rank regularization. 'Block-sparse' attribute is also preserved by sparse representation with respect to the recovered dictionary. We construct 3 independent subspaces $\{\mathbf{S}_i\}_{i=1}^3 \subset \mathbb{R}^{100}$, the rank of each base $\mathbf{U}_i \in \mathbb{R}^{100\times4}$ is 4, so each subspace has a dimension of 4. We obtain the contaminated data matrix $\mathbf{D} \in \mathbb{R}^{100\times600}$ by sampling 200 data from each subspace and adding Gaussian noise to the generated data. RSSC problem is solved by inexact ALM as detailed in Algorithm 1.

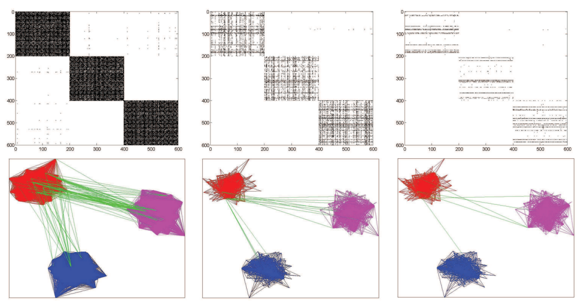

Figure 3: An illustration of sparse presentation with different iterations. Top: Sparse coefficients after 10, 20, 40 iterations. Bottom: Similarity graphs after 10, 20, 40 iterations.

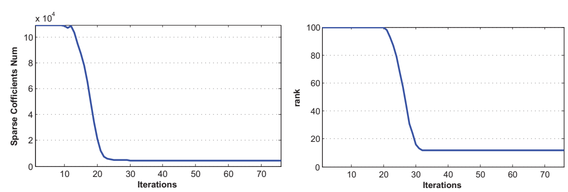

Figure 4: Left:The non-zero number of coefficients in sparse representation matrix as iteration continues. Left:The rank of recovered matrix as iteration continues.

Fig.3 shows the sparse coefficients and corresponding similarity graphs after 10, 20, 40 iterations, the representation is more sparser as the iterations continues. Fig.4 shows the quantitative result of the changes of non-zero number of sparse coefficients and the rank of the recovered dictionary as iteration continues. The non-zero coefficients number reduces from 109,572 to 4,360, the rank of data matrix reduces from 100 to 12.

Table 1: Average errors(%) for sequences with 2 motions.

Two Motions					
	GPCA	LSA	RANSAC	SSC	RSSC
Checkerboard: 78 sequences					
Average	6.45	8.37	5.62	**1.41**	1.67
Traffic: 31 sequences					
Average	2.33	4.75	2.07	1.95	**1.39**
Articulated: 11 sequences					
Average	**0.79**	3.29	1.82	3.70	0.81
All: 120 sequences					
Average	3.19	5.47	3.17	2.35	**1.29**
Three Motions					
	GPCA	LSA	RANSAC	SSC	RSSC
Checkerboard: 26 sequences					
Average	31.68	30.25	25.43	6.04	**3.52**
Traffic: 7 sequences					
Average	26.88	22.93	16.06	7.57	**3.01**
Articulated: 2 sequences					
Average	31.48	23.80	9.39	6.34	**3.24**
All: 35 sequences					
Average	30.01	25.66	16.96	6.65	**3.26**

3.2 Real Data: the Hopkins155 Database

In this section, we evaluate our RSSC approach compared to other state-of-the-art methods on the Hopkins 155 motion database [9], such as GPCA [4], LSA [10], RANSAC [2] and SSC [5]. The database consists of 155 video sequences, each of which has two or three motions which can be divided into three categories: checkerboard, traffic and articulated sequences. Some example frames are illustrated in Fig.1. The points trajectories are extracted automatically by a tracker and the trajectories are always contaminated by noise. Table 1 shows detailed average errors for 120 sequences (78 checkerboard, 31 traffic, 11 articulated sequences) with two motions and 35 sequences (26 checkerboard, 7 traffic, 2 articulated sequences) with three motions. Obviously, our RSSC method performs best among all algorithms with lowest average errors.

4. Conclusions

In this paper, we propose a convex regularization based subspace clustering method for robust motion segmentation. By introducing low-rank and sparse regularizations, our RSSC simultaneously removes noise from data and performs robust sparse representation. The experimental results demonstrate the superiority and robustness of our method over other state-of-the-art methods.

Acknowledgement

This work is supported by the National Natural Science Foundation of China (90920001 and 61101212), the National High Technology Research and Development Program of China (863 Program) No. 2012AA012505, the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, and the Fundamental Research Funds for the Central Universities.

References

[1] J. Park, H. Zha, and R. Kasturi, “Spectral clustering for robust motion segmentation,” *ECCV*, pp. 1 – 8, 2004.

[2] M. A. Fischler and R. C. Bolles, “Ransac random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, pp. 381 – 395, 1981.

[3] K. Kanatani, “Motion segmentation by subspace separation and model selection,” *ICCV*, pp. 586 – 591, 2001.

[4] R. Vidal, Y. Ma, and S. Sastry, “Generalized principal component analysis (gpca),” *TPAMI*, pp. 1 – 15, 2005.

[5] E. Elhamifar and R. Vidal, “Sparse subspace clustering,” *CVPR*, pp. 2790 – 2797, 2009.

[6] Z. Lin, M. Chen, L. Wu, and Y. Ma, “The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices,” *UIUC Technical Report*, 2009.

[7] D. L. Donoho, “For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution,” *Communications on Pure and Applied Mathematics*, pp. 797 – 829, 2006.

[8] E. J. Candes and B. Recht, “Exact matrix completion via convex optimization,” *Foundations of Computational Mathematics*, pp. 717 – 772, 2009.

[9] R. Tron and R. Vidal, “A benchmark for the comparsion of 3-d motion segmentation algorithms,” *CVPR*, pp. 1 – 8, 2007.

[10] J. Yan and M. Pollefeys, “A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate,” *ECCV*, pp. 94 – 106, 2006.