Neurocomputing 198 (2016) 109–113

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

An eigen decomposition based rank parameter selection approach for the NRSFM algorithm

Yang Liu [a], Zhan-Li Sun [a,*], Ya-Ping Wang [a], Li Shang [b]

[a] School of Electrical Engineering and Automation, Anhui University, Hefei, China
[b] Department of Communication Technology, Electronic Information Engineering College, Suzhou Vocational University, Suzhou, China

ARTICLE INFO

Article history:
Received 5 March 2015
Received in revised form
30 May 2015
Accepted 20 June 2015
Available online 11 March 2016

Keywords:
Non-rigid structure from motion
Rank parameter selection
Eigen decomposition

ABSTRACT

Non-rigid structure from motion (NRSFM) with an affine structure from motion (aSFM) kernel (NRSFM-aSFM) is a relative novel and robust 3D shape recovery algorithm to the abrupt deformations. Nevertheless, the estimated 3D shapes generally fluctuate with the variation of rank parameter, i.e., the number of shape bases. Therefore, it is necessary to develop an effective method to select the rank parameter. In this paper, we propose an eigen decomposition based rank parameter selection approach, which can automatically select the optimal or an approximately optimal rank parameter for the NRSFM-aSFM algorithm. In the proposed method, a symmetric matrix is first constructed to simplify the singular value estimation. Further, the QR decomposition based iterations are carried out to compute the eigenvalues of the observation matrix. Finally, the rank parameter is estimated according to the cumulative sum of eigenvalues by setting a referred threshold value. The experimental results on several widely used sequences demonstrate the effectiveness and feasibility of the proposed method.

1. Introduction

The technique of non-rigid structure from motion (NRSFM) provides an approach to jointly estimate 3D object shapes and the relative camera motions from the corresponding 2D points in a sequence of images [1,2]. Generally, the estimated 3D information can effectively enhance the capabilities of existing image processing systems [3–7]. Nevertheless, during the process of movement, the objects generally undergo a series of shape deformations and pose variations [1]. Due to the under-constrained nature and the absence of the prior knowledge on 3D shape, how to achieve good solutions of NRSFM is still a very difficult and a more challenging task in computer vision [8–10].

In order to make NRSFM more tractable, i.e., compress the deformation model and reduce the number of unknowns to estimate, 3D shape deformation is usually constrained to be smooth over time [11–14]. Nevertheless, the temporal smoothness cannot be enforced when the data lacks temporal ordering. Moreover, the benefits brought by the temporal smoothness constraint are less evident when objects undergo the abrupt deformations.

To address this problem, instead of the temporal constraint, a relative novel and robust NRSFM algorithm with an affine structure from motion (aSFM) kernel (NRSFM-aSFM) is proposed in [2] by enforcing spatial smoothness. A prominent advantage of the approach is that it can effectively deal with the data lacking temporal ordering or with abrupt deformations. For the NRSFM-aSFM algorithm, one problem is that the reconstruction accuracies generally fluctuate with the variation of rank parameter K, i.e., the number of shape bases. Generally, the parameter is experimentally determined, i.e., different K values are tried one by one. Nevertheless, the parameter derived from one image sequence may be not a good choice for other image sequences. Therefore, it is necessary to design an effective method to automatically select the optimal or an approximately optimal value for the rank parameter.

In this paper, an eigen decomposition based rank parameter selection approach is proposed for the NRSFM-aSFM algorithm. In the proposed method, the rank parameter is estimated according to the cumulative sum of eigenvalues, which are obtained via the QR decomposition. By means of the proposed method, the optimal or an approximately optimal K can be automatically derived from the 2D tracking data. The experimental results on several widely used sequences verify the effectiveness and feasibility of the proposed approach.

* Corresponding author.
E-mail address: zhlsun2006@126.com (Z.-L. Sun).

http://dx.doi.org/10.1016/j.neucom.2015.06.110
0925-2312/

The remainder of the paper is organized as follows. The proposed method is presented in Section 2. Experimental results are given in Section 3. Finally, conclusions are made in Section 4.

2. Methodology

Assume that $[x_{t,j}, y_{t,j}]^T$ $(t = 1, 2, \ldots, T, j = 1, 2, \ldots, n)$ is the 2D projection of the jth 3D point observed on the tth image, i.e., x and y coordinates of feature points. The n input 2D point tracks of T images can be represented as a $2T \times n$ observation matrix $\mathbf{W}$, i.e.,

$$\mathbf{W} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ y_{1,1} & y_{1,2} & \cdots & y_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,n} \\ y_{T,1} & y_{T,2} & \cdots & y_{T,n} \end{pmatrix}. \tag{1}$$

According to the linear subspace model, $\mathbf{W}$ can be factorized as [1]:

$$\mathbf{W} = \mathbf{MS} = \mathbf{D}(\mathbf{C} \otimes \mathbf{I}_3)\mathbf{S}, \tag{2}$$

where $\mathbf{M}$ is a motion factor that is composed of affine transformation matrices of each frame, $\mathbf{S}$ denotes a shape matrix including K shape bases. The matrices $\mathbf{D}$, $\mathbf{C}$ and $\mathbf{I}_3$ represent a block-diagonal rotation matrix, a shape coefficient matrix and a 3×3 identity matrix, respectively. The expression $\mathbf{C} \otimes \mathbf{I}_3$ denotes the Kronecker product of $\mathbf{C}$ and $\mathbf{I}_3$.

Let $\mathbf{c}_t^T$ the tth row of $\mathbf{C}$, the 3D shape of tth image can be modeled as a linear combination of K shape bases $\mathbf{S}_k \in \mathbb{R}^{3 \times n}$ [1], i.e.,

$$S(\mathbf{c}_t^T) = (\mathbf{c}_t^T \otimes \mathbf{I}_3)\mathbf{S} = \sum_{k=1}^{K} c_{t,k}\mathbf{S}_k, \tag{3}$$

For this linear model, we can understand that $\mathbf{S}$ represents a linear shape space, and $\mathbf{c}_t^T$ is the corresponding coordinates of 3D shape bases $\mathbf{S}_k$.

It can be seen from (3) that, in order to recover the 3D shape of tth image, we should first determine the rank parameter K, i.e., the number of shape bases. In our proposed method, K is estimated according to the cumulative sum of singular values $\lambda_i (i = 1, \ldots, K)$ of $\mathbf{W}$. Let $\mathbf{A} = \mathbf{WW}'$, the singular values λ_i can be estimated as the eigenvalues of $\mathbf{WW}'$ via the following optimization model:

$$\| \mathbf{A} - \lambda_i^2 \mathbf{I} \| = 0. \tag{4}$$

In order to obtain the solutions λ_i of (4), $\mathbf{A}$ is first transformed into an upper Heisenberg matrix $\mathbf{A}_1$ via the Householder transformation. Furthermore, $\mathbf{A}_1$ is factorized as a product of two matrices $\mathbf{Q}_1$ and $\mathbf{R}_1$ via the **QR** decomposition [15],

$$\mathbf{A}_1 = \mathbf{Q}_1\mathbf{R}_1.$$

Subsequently, the following iterations are carried out,

$$\mathbf{A}_2 = \mathbf{R}_1\mathbf{Q}_1 = \mathbf{Q}_1^T\mathbf{A}_1\mathbf{Q}_1, \tag{5}$$

$$\vdots$$

$$\mathbf{A}_k = \mathbf{R}_{k-1}\mathbf{Q}_{k-1} = \mathbf{Q}_{k-1}^T\mathbf{A}_{k-1}\mathbf{Q}_{k-1}, \tag{6}$$

$$\mathbf{A}_{k+1} = \mathbf{R}_k\mathbf{Q}_k = \mathbf{Q}_k^T\mathbf{A}_k\mathbf{Q}_k, \tag{7}$$

till converges to an upper triangular matrix,

$$\mathbf{A}_k \to \mathbf{R} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ & \lambda_2 & \cdots & * \\ & & \ddots & \vdots \\ & & & \lambda_n \end{pmatrix}. \tag{8}$$

Specifically, if $\mathbf{A}_k$ is a symmetric matrix, $\mathbf{A}_k$ will converge to an approximately diagonal matrix, i.e.,

$$\mathbf{A}_k \to \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}. \tag{9}$$

Given a threshold η, according to the percentage of the cumulative sum of $\lambda_i (\lambda_1 > \lambda_2 > \cdots > \lambda_n)$, K is estimated as the smallest number that fulfills

$$K = \min\left\{ s \in N \mid \frac{\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_s^2}{\sum_{i=1}^{n} \lambda_i^2} \geq \eta \right\}, \tag{10}$$

where the parameter η is set to be a value so that most nonzero singular values are included.

After determining K, we then compute $\mathbf{M}$ and $\mathbf{S}$ of (2). For two observations $\mathbf{w}^t$ and $\mathbf{w}^{t'}$, the affine structure from motion (aSFM) rotation invariant kernel (RIK) is defined as [2]:

$$\kappa(\mathbf{w}^t, \mathbf{w}^{t'}) = \exp\left(\frac{-r_{t,t'}^2}{\sigma^2} \right) + \alpha\delta_{t,t'}, \tag{11}$$

where σ, α and $r_{t,t'}^2$ are the kernel scale, the regulation parameter, and the reprojection error, respectively. According to (11), we can first obtain a complete kernel matrix $\mathbf{Kww}$ of $\mathbf{W}$, and then compute its eigenvector matrix $\mathbf{V}$ associated with the d largest eigenvalues in the diagonal matrix Λ. The matrix $\mathbf{M}$ of (2) can be computed as [2]:

$$\mathbf{M} = \mathbf{D}(\mathbf{K_{WW}}\mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{X} \otimes \mathbf{I}_3), \tag{12}$$

where $\mathbf{D}$ and the coefficient matrix $\mathbf{X}$ are obtained by the metric upgrade algorithm [13]. According to (2), the matrix $\mathbf{S}$ can be given by:

$$\mathbf{S} = \mathbf{M}^{\dagger}\mathbf{W}, \tag{13}$$

where $\mathbf{M}^{\dagger}$ is the Moore–Penrose pseudo-inverse of $\mathbf{M}$. Further,

$$\mathbf{c}_t^T = f(\mathbf{w}^t) = \boldsymbol{\kappa}(\mathbf{w}^t, \mathbf{W})\mathbf{V}\mathbf{\Lambda}^{-1/2}\mathbf{X}. \tag{14}$$

Finally, in terms of (2), (13), (14), and the obtained rank parameter K, the 3D shape of the tth image can be estimated as [2]:

$$S(\mathbf{c}_t^T) = (\mathbf{c}_t^T \otimes \mathbf{I}_3)\mathbf{M}^{\dagger}\mathbf{W} \tag{15}$$

3. Experiments

3.1. Experimental data and set-up

We evaluate the performance of our proposed method on seven non-rigid motion datasets. The seven motion capture sequences are: *walking, face1, stretch, dance, face2, jaws, yoga*, respectively. For these sequences, the corresponding number of frames (T) and the number of point tracks (n) are given in Table 1. As an example, Fig. 1 shows one frame of the seven image sequences.

Table 1
The number of frames (T) and the number of point tracks (n) for seven motion capture sequences.

Dataset	T	n
walking	260	55
face1	74	37
stretch	370	41
dance	264	75
face2	316	40
jaws	240	91
yoga	307	41