Joint SVD and Its Application to Factorization Method

Gen Hori[1,2]

[1] Faculty of Business Administration, Asia University, Tokyo 180-8629, Japan
[2] Lab. for Advanced Brain Signal Processing, BSI, RIKEN, Saitama 351-0198, Japan
hori@brain.riken.jp

Abstract. This paper introduces an application of joint SVD to the factorization method which is a standard method in computer vision for the estimation of the camera motion and the object shape from an image stream taken by a camera that moves around the object. For computer vision systems with several cameras installed for the same direction, we implement a new algorithm for estimation of camera motion matrix that utilizes the measurement matrices from all the cameras based on joint SVD.

1 Introduction

The factorization method proposed by Tomasi and Kanade[1] is a method for the estimation of the camera motion and the object shape from an image stream taken by a camera that moves around the object. While almost previous methods for the same purpose are weak for noise and give unstable solution, the factorization method has a large advantage in that it is robust under noisy situation and gives a stable solution because it is based on well-established SVD (singular value decomposition).

On the other hand, joint SVD is a problem of finding a pair of unitary matrices which simultaneously diagonalizes several (possibly non-square) matrices. Joint SVD problem appears as a natural extension of joint diagonalization problem which has applications in signal separation methods and structured eigenvalue problems. Pesquet-Popescu et al.[2] introduced a Jacobi-like algorithm for joint SVD problem and proposed to apply joint SVD to obtaining separable representation of images. Hori[4] proposed a pair of matrix gradient flows to solve joint SVD problem.

This paper proposes to apply joint SVD to the estimation of the camera motion matrix for computer vision systems with several cameras installed for the same direction. The proposed method utilizes the measurement matrices from all the cameras based on joint SVD. A preliminary experiment using a computer-generated measurement matrix shows that the proposed method improves the estimation of the camera motion matrix comparing to the factorization method based on the standard SVD.

V. Vigneron et al. (Eds.): LVA/ICA 2010, LNCS 6365, pp. 563–570, 2010.

2 Factorization Method

This section summarizes the factorization method introduced by Tomasi and Kanade[1] which recovers the camera motion and the object shape from the image stream taken by the camera that moves around the object. For simplicity, we restrict our attention to the orthographic projection model (in which the three dimensional space is orthogonally projected to the image plane) throughout the study.

Suppose that we have an image stream which consists of F images taken by the camera that moves around the object. From the first image of the stream, we find P feature points on the object and track them throughout the F images to obtain the coordinates

$$\{(x_{fp}, y_{fp}) \mid f = 1, \ldots, F, p = 1, \ldots, P\}.$$

From the coordinates, we define a $2F \times P$ matrix

$$\tilde{A} = \begin{pmatrix} x_{11} & \cdots & x_{1P} \\ \vdots & & \vdots \\ x_{F1} & \cdots & x_{FP} \\ y_{11} & \cdots & y_{1P} \\ \vdots & & \vdots \\ y_{F1} & \cdots & y_{FP} \end{pmatrix}.$$

Next, we calculate the row-wise averages of the matrix $\tilde{A}$,

$$\bar{x}_f = \frac{1}{P} \Sigma_{p=1}^{P} x_{fp}, \quad \bar{y}_f = \frac{1}{P} \Sigma_{p=1}^{P} y_{fp},$$

and subtract them from the corresponding rows of $\tilde{A}$ to define a new matrix A

$$A = \tilde{A} - \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_F \\ \bar{y}_1 \\ \vdots \\ \bar{y}_F \end{pmatrix} \begin{pmatrix} 1 & \cdots & 1 \end{pmatrix}$$

whose rows are zero-mean. The matrix A is called "measurement matrix" and consists of the relative coordinates of the feature points to the centroid,

$$(x_{fp} - \bar{x}_f, y_{fp} - \bar{y}_f).$$

The measurement matrix A is factorized into two matrices M and S that correspond to the camera motion and the object shape respectively. As described in

Fig.1, we denote the coordinate of the camera and two unit vectors that define the camera direction for the f-th image by

$$t_f = (t_{fx}, t_{fy}, t_{fz})^T, \quad i_f = (i_{fx}, i_{fy}, i_{fz})^T, \quad j_f = (j_{fx}, j_{fy}, j_{fz})^T$$

where ^{T} denotes the transpose and the coordinate of the p-th feature point in the three dimensional space by

$$s_p = (s_{px}, s_{py}, s_{pz})^T,$$

from which we define a $2F \times 3$ matrix M that corresponds to the camera motion

$$M = \begin{pmatrix} {i_1}^T \\ \vdots \\ {i_F}^T \\ {j_1}^T \\ \vdots \\ {j_F}^T \end{pmatrix} = \begin{pmatrix} i_{1x} & i_{1y} & i_{1z} \\ \vdots & \vdots & \vdots \\ i_{Fx} & i_{Fy} & i_{Fz} \\ j_{1x} & j_{1y} & j_{1z} \\ \vdots & \vdots & \vdots \\ j_{Fx} & j_{Fy} & j_{Fz} \end{pmatrix},$$

and a $3 \times P$ matrix S that corresponds to the object shape

$$S = \begin{pmatrix} s_1 & \cdots & \cdots & s_P \end{pmatrix} = \begin{pmatrix} s_{1x} & \cdots & \cdots & s_{Px} \\ s_{1y} & \cdots & \cdots & s_{Py} \\ s_{1z} & \cdots & \cdots & s_{Pz} \end{pmatrix}.$$

Here we set the origin of the three dimensional space to the centroid of the feature points so that the rows of S are zero-mean,

$$\textstyle\sum_{p=1}^{P} s_p = 0.$$

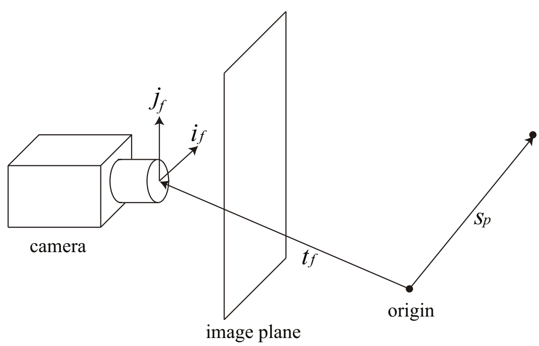

Fig. 1. Coordinate system

Because the coordinate of the p-th feature point on the f-th image is given as

$$x_{fp} = {i_f}^T (s_p - t_f), \quad y_{fp} = {j_f}^T (s_p - t_f),$$

and the rows of A and S are zero-mean, A is factorized as

$$A = MS \tag{1}$$

not depending on t_f. This leads to the fact that the rank of the measurement matrix A is 3 when it is not influenced by any noise. In practice of the factorization method, we obtain the decomposition (1) using the standard SVD (singular value decomposition) as follows. The first step is to decompose the $2F \times P$ measurement matrix A by SVD as

$$A = \tilde{U}\tilde{\Sigma}\tilde{V}^T$$

where $\tilde{U}$ is a $2F \times P$ matrix, $\tilde{\Sigma}$ a $P \times P$ diagonal matrix and $\tilde{V}$ a $P \times P$ matrix. Even when the measurement matrix A is influence by some noise, it is almost rank-3, that is, the diagonal elements of $\tilde{\Sigma}$ other than the dominant three are negligible so that we can truncate $\tilde{U}$ to a $2F \times 3$ matrix U, $\tilde{\Sigma}$ to a 3×3 matrix Σ and $\tilde{V}$ to a $P \times 3$ matrix V to have

$$A = U\Sigma V^T.$$

The second step is to find an invertible matrix T that gives the solution as

$$\begin{aligned} M &= UT \\ S &= T^{-1}\Sigma V^T \end{aligned}$$

using the fact that the rows of the matrix M need to meet the following conditions

$${i_f}^T i_f = 1, \quad {j_f}^T j_f = 1, \quad {i_f}^T j_f = 0 \qquad (f = 1, \ldots, F). \tag{2}$$

It is shown that the conditions (2) give a unique solution for the matrix T therefore the unique M and S up to the ambiguity related to the choice of the coordinate system.

3 Joint SVD

Given a set of K real or complex $m \times n$ matrices

$$\{A_1, A_2, \ldots, A_K\},$$

joint SVD or simultaneous SVD is a problem of finding orthogonal or unitary matrices U and V which make

$$\{U^* A_1 V, U^* A_2 V, \ldots, U^* A_K V\}$$

as diagonal as possible simultaneously. Joint SVD problem appears as a natural extension of joint diagonalization problem which has applications in signal separation methods and structured eigenvalue problems. We say that a joint SVD problem is *left-exact* (*right-exact*) when all the given matrices share a common set of left (right) singular vectors. We say that the problem is *exact* when it is *left-exact* and *right-exact* and the left and right singular vectors have one-to-one correspondence over the given matrices. From a purely mathematical viewpoint, a joint SVD problem can be solved exactly only if the problem is *exact*, where the SVD of an arbitrary single matrix A_k gives the solution to the problem. From a practical viewpoint, however, the problem is rarely *exact* and we have to find optimal solution with respect to some diagonality criterion. One of the typical situations in practical applications is that additive noise on observed matrices makes the problem *non-exact* although the problem is theoretically *exact*.

Hori[4] proposed to use the gradient ascent equations of the potential function

$$\varphi(U, V) = \sum_{k=1}^{K} \phi_k(U^* A_k V) \tag{3}$$

to solve a joint SVD problem where each $\phi_k(A)$ is a diagonality criterion supposed to take its maximum when A is an extended diagonal matrix. The general gradient ascent equations are derived as

$$\begin{aligned}\frac{dU}{dt} &= \frac{1}{2}U\sum_{k=1}^{K}((U^*A_kV)(\frac{d\,\phi_k}{dA}(U^*A_kV))^* \\ &\qquad -(\frac{d\,\phi_k}{dA}(U^*A_kV))(U^*A_kV)^*),\end{aligned} \tag{4}$$

$$\begin{aligned}\frac{dV}{dt} &= \frac{1}{2}V\sum_{k=1}^{K}((U^*A_kV)^*(\frac{d\,\phi_k}{dA}(U^*A_kV)) \\ &\qquad -(\frac{d\,\phi_k}{dA}(U^*A_kV))^*(U^*A_kV)).\end{aligned} \tag{5}$$

For example, we define the following diagonality criterion

$$\phi_k(A) = \sum_{1\le j\le n} |a_{jj}|^2, \qquad k = 1, \ldots, K,$$

which takes its maximum when A is an extended diagonal matrix, and substitute it in (4) and (5) to obtain a pair of gradient flows for joint SVD,

$$\begin{aligned}\dot{U} &= U\sum_{k=1}^{K}((U^*A_kV)\mathrm{diag}(U^*A_kV)^* \\ &\qquad -\mathrm{diag}(U^*A_kV)(U^*A_kV)^*),\end{aligned} \tag{6}$$

$$\begin{aligned}\dot{V} &= V\sum_{k=1}^{K}((U^*A_kV)^*\mathrm{diag}(U^*A_kV) \\ &\qquad -\mathrm{diag}(U^*A_kV)^*(U^*A_kV)).\end{aligned} \tag{7}$$

There are other possible approaches to the joint SVD problem. For example, consider two joint diagonalization problems of Hermitian matrices,

$$\{A_1 A_1{}^*, A_2 A_2{}^*, \ldots, A_K A_K{}^*\}$$

and

$$\{A_1{}^* A_1, A_2{}^* A_2, \ldots, A_K{}^* A_K\},$$

and suppose that U and V are the solutions to the respective problems obtained by some standard joint diagonalization procedure, then

$$\{U^* A_1 V, U^* A_2 V, \ldots, U^* A_K V\}$$

is expected to be the joint SVD of the given matrices. Hori[3] compares the approach to joint SVD via joint diagonalization and the direct approach to join SVD to conclude that the direct approach gives more accurate solution.

4 Factorization Method for Several Cameras

In certain kinds of computer vision systems including two-eyed robot vision systems, several cameras are installed for the almost same direction. This section considers the factorization method for such systems in which K cameras are installed. Fig.2 describes the case of $K = 2$. We consider the case where the relative positions of the cameras are invariant throughout the image stream and all the cameras are set to the same direction, that is,

$$\begin{array}{l} i_f{}^{(1)} = i_f{}^{(2)} = \cdots = i_f{}^{(K)} \\ j_f{}^{(1)} = j_f{}^{(2)} = \cdots = j_f{}^{(K)} \end{array} \quad (f = 1, \ldots, F).$$

In this case, the camera motion matrix M should be common to all the cameras. However each camera gives each measurement matrix A_k which is subject to noise and the camera-wise SVD of the measurement matrix

$$A_k = U_k \Sigma_k V_k{}^T \ (k = 1, \ldots, K)$$

possibly results in different U_k therefore different camera motion matrix M_k. For this situation, we propose to consider the real joint SVD of $A_1, \ldots, A_K$ which makes

$$U^T A_k V \ (k = 1, \ldots, K)$$

as diagonal as possible simultaneously. This gives a common U and therefore a common camera motion matrix M and the resulting camera motion matrix is expected to be more accurate than the one obtained from the SVD of a single measurement matrix of some representative camera because the proposed method utilizes the information from all the cameras.

Fig.3 shows the result of a preliminary experiment with a computer-generated 250×125 measurement matrix $A = MS$ where the camera motion matrix M is generated by changing the yaw, pitch and roll angles from $-10°$ to $10°$ and the

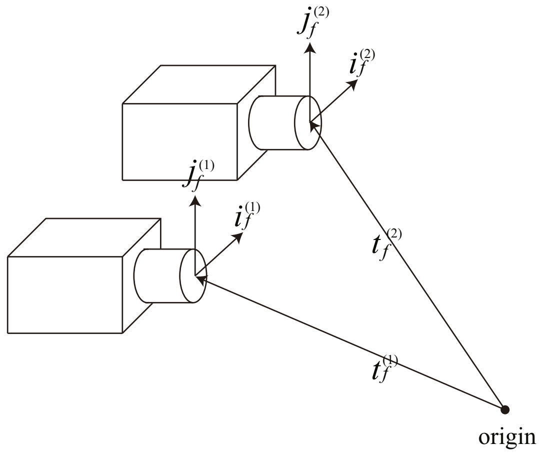

Fig. 2. Parallel cameras

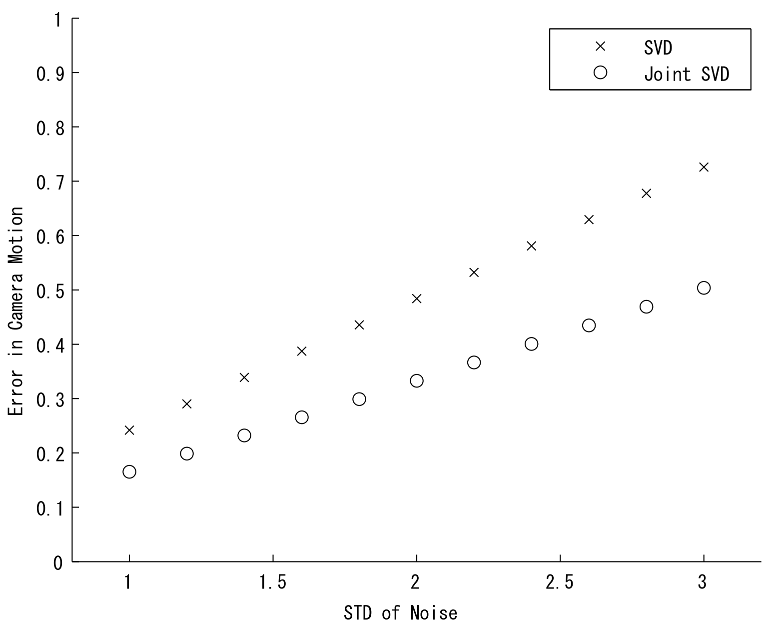

Fig. 3. Preliminary experiment

object shape matrix S is of 125 equally-spaced lattice points in the three dimensional space. We suppose that the whole object is projected onto a quarter of the image plane of 1000×1000 pixels and change the standard deviation of the Gaussian noise added to the elements of the measurement matrix A from 1 pixel to 3 pixels. The camera motion matrix is estimated using the factorization method based on the standard SVD of a single noisy measurement matrix and joint SVD of two noisy measurement matrices. The difference between the true camera motion matrix and the estimated one is measured using the matrix Frobenius norm and averaged over 10 trials. From the result, we see that the proposed method based on the joint SVD of two measurement matrices improves the estimation of the camera motion matrix comparing to the factorization method based on the SVD of a single measurement matrix.

5 Concluding Remarks

We have proposed to apply joint SVD to the estimation of the camera motion matrix in the computer vision systems with several cameras installed for the same direction. A preliminary experiment using a computer-generated measurement matrix has shown that the proposed method improves the estimation of the camera motion matrix comparing to the factorization method based on the standard SVD. Our future study includes the extension of the proposed method to more precise projection models such as the scaled orthographic projection model and the paraperspective projection model as well as the experiments using real image stream data taken by parallel cameras.

References

1. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography - A factorization method. Int. J. Comput. Vision 9-2, 137–154 (1992) Technical Report CMU-CS-91-172, CMU (1991)
2. Pesquet-Popescu, B., Pesquet, J.-C., Petroplu, A.P.: Joint singular value decomposition - a new tool for separable representation of images. In: Proc. 2001 Intl. Conf. on Image Processing, vol. 2, pp. 569–572 (2001)
3. Hori, G.: Comparison of Two Main Approaches to Joint SVD. In: Proc. Intl. Workshop Independent Component Analysis Blind Signal Separation, pp. 42–49 (2009)
4. Hori, G.: A general framework for SVD flows and joint SVD flows. In: Proc. IEEE Intl. Conf. Acoustics Speech Signal Processing, vol. 2, pp. 693–696 (2003)
5. Hori, G.: A new approach to joint diagonalization. In: Proc. Intl. Workshop Independent Component Analysis Blind Signal Separation, pp. 151–155 (2000)
6. Hori, G., Tanaka, T.: Finding initial values for time-varying joint diagonalization. In: Proc. IEEE Intl. Conf. Acoustics and Speech Signal Processing, pp. 2018–2021 (2010)
7. Cardoso, J.-F., Souloumiac, A.: Blind beamforming for non Gaussian signals. IEEE Proc.-F 140, 362–370 (1993)
8. Yeredor, A.: Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. IEEE Trans. Signal Processing 50(7), 1545–1553 (2002)
9. Pham, D.T.: Joint approximate diagonalization of positive definite Hermitian matrices. SIAM J. Matrix Anal. Appl. 22, 1136–1152 (2001)
10. Ziehe, A., Laskov, P., Nolte, G., Müller, K.-R.: A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. J. Machine Learning Research 5, 777–800 (2004)