Robust Affine Motion Estimation in Joint Image Space using Tensor Voting*

Eun-Young Kang, Isaac Cohen and Gérard Medioni
Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California 90089-0273
{elkang, icohen, medioni}@iris.usc.edu

Abstract

*Robustness of parameter estimation relies on discriminating inliers from outliers within the set of correspondences. In this paper, we present a method using tensor voting to eliminate outliers and estimating affine transformation parameters directly from covariance matrix of selected inliers without additional parameter estimation processing. Our approach is based on the representation of the correspondences in a decoupled joint image space and the use of the metric associated with the affine transformation. We enforce the metric property in a joint image space for tensor voting, detect several inlier groups corresponding distinct affine motions and directly estimate affine parameters from each set of inliers. The proposed approach is illustrated by a set of challenging examples.**

1. Introduction

Motion estimation using parametric models is widely used for video processing such as image mosaics, video compression and video surveillances [3][4][6][8]. The affine motion model is a commonly used method for these applications due to its simplicity and the small inter-frame camera motion. In this paper we present a robust and non-iterative correspondence-based method to estimate affine parameters using tensor voting.

Robustness of parameter estimation depends on successfully removing the outliers within correspondences. Several techniques were proposed for extracting good correspondences that fit the targeted parametric model [7]. RANSAC[9] and its enhanced variations are commonly used techniques because of the capability of handling a large portion of outliers. However, these techniques perform iterative steps that are sensitive to the selections of samples and thresholds, and they do not constrain the space of admissible solutions according to the parametric model used. Recently non-iterative tensor voting based-methods

for estimating fundamental matrix were proposed in [1][5]. The tensor voting formalism was used as a pre-processing step for outlier removal in 8D and 4D space by characterizing hyper-surfaces. In 8D approach, a plane is parameterized by 8 fundamental matrix variables, and the outliers are the points not on the plane. In this approach, a local smoothness and a global hyper-plane constraint are used at the same time. However, the parameters defining the 8D basis make the space neither orthogonal nor isotropic. Therefore, the input must be properly scaled prior to processing. In [5], a 4D approach using the joint image space is derived from point correspondences to reduce the dimensionality and to provide isotropic and orthogonal properties. Inliers are detected as points on a 4D cone as defined by the epipolar constraint in 4D joint image space.

In this paper, taking advantages of tensor voting-based methods [1][5], we propose an affine motion estimation method using tensor voting to detect inliers and outliers and recover corresponding parameters. The proposed method defines a decoupled joint image space from input correspondences, shows that affine transformation constraint represents a 2D plane in the defined space and enforces 2D plane structure for tensor voting to detect inliers. Additionally, our method detects several independent affine motions and directly estimates the parameters from each set of inliers.

In section 2, we briefly describe the tensor voting method [2]. The following section describes affine transformation properties in the joint image space and characterizes a decoupled joint image space and the corresponding metric. It allows us to define direct parameter estimation through the tensor voting process. Finally, a set of experimental results illustrates our methodology performed on video data sets and compares these results to the RANSAC algorithm.

2. Tensor Voting

Tensor voting formalism provides a robust approach for extracting salient structures by encoding data and the corresponding uncertainties in a second order symmetric tensor. An efficient voting process allowing for the

propagation of local properties complements this data representation. The extraction of salient structures is inferred from the canonical description of an arbitrary tensor by its eigensystem representing the local geometric properties of the data. Indeed, any arbitrary symmetric tensor can be decomposed by:

$$S = (\lambda_1 - \lambda_2)e_1e_1^T + \sum_{i=1}^{n-1}(\lambda_i - \lambda_{i+1})\sum_{j=1}^{i} e_je_j^T + \lambda_n \sum_{i=1}^{n} e_ie_i^T$$

where λ_i denote the eigenvalues (sorted in a decreasing order) and e_i denotes corresponding eigenvectors. In any dimension higher than 3D, the first term of S characterizes the hyper-plane orientation (normal) and the associated ($\lambda_1 - \lambda_2$) saliency. These local geometric properties are propagated within a domain of influence depending the principal orientation (given by e_1) and on the associated saliency.

3. Affine Model and Tensor Voting

In this section, we show that a decoupled joint image space is a 4D space, and that the embedded structure that represents the affine motion is a 2D plane. By inferring the most salient 2D plane from input correspondences based on tensor voting, we remove the outliers and estimate the parameters directly from inliers. This approach formulates the problem in a geometric space and minimizes geometric distance in a non-iterative manner. It differs from classical techniques, in that they attempt to minimize the algebraic errors iteratively.

3.1. Affine Model and Joint Image Space

A 2D affine model is defined by six parameters. In this model, a correspondence between a feature point (x,y) from one image and the corresponding point (x',y') from the second image is given by the following equation:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix}$$

which can be rewritten in the parametric space as:

$$\begin{pmatrix} x & y & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & x & y & 1 \end{pmatrix}\begin{pmatrix} a & b & t_x & c & d & t_y \end{pmatrix}^T = \begin{pmatrix} x' \\ y' \end{pmatrix}$$

A set of linear equations derived from corresponding points are usually solved by a least square method or its variations that minimize algebraic errors. In the traditional joint image space representation, each point is a combination of 2D image vectors and affine transformation can be rewritten as: $(q^T 1)P^T P \begin{pmatrix} q \\ 1 \end{pmatrix} = 0$

where $q = (x,y,x',y',1)$ and $P = \begin{pmatrix} a & b & -1 & 0 & t_x \\ c & d & 0 & -1 & t_y \end{pmatrix}^T$

But in the affine model, the joint spaces (x,y,x') and (x,y,y') are independently constrained and therefore can be decoupled to reduce the dimension of the joint image space.

Therefore, by defining $p_x = (a,b,-1,t_x)^T$ and $p_y = (c,d,-1,t_y)^T$, we have two separate joint spaces $q_x = (x,y,x',1)^T$ and $q_y = (x,y,y',1)^T$. We obtain the following equations in the decoupled joint image spaces: $q_x^T C_x q_x = 0$ and $q_y^T C_y q_y = 0$ where C_1 and C_2 are defined by:

$$C_x = p_x^T p_x = \begin{pmatrix} a^2 & ab & -a & at_x \\ ab & b^2 & -b & bt_x \\ -a & -b & 1 & -t_x \\ at_x & bt_x & -t_x & t_x^2 \end{pmatrix} \quad C_y = p_y^T p_y = \begin{pmatrix} c^2 & cd & -c & ct_y \\ cd & d^2 & -d & dt_y \\ -c & -d & 1 & -t_y \\ ct_y & dt_y & -t_y & t_y^2 \end{pmatrix}$$

In this representation, each 4D point defined by $q_x = (x,y,x',1)^T$ lies on a 2D plane parameterized by $p_x = (a,b,-1,t_x)^T$ in the 4D space. Therefore, the points on the 2D plane define inliers. In case that input correspondences consist of perfect correspondences, the eigenvector corresponding to the smallest eigenvalue of the covariance matrix of the correspondences in the space *(x,y,x',1) or (x,y,y',1)* characterizes the parameters of the 2D plane. If several affine motions are present among the correspondences, the same number of corresponding 2D planes is defined in the 4D space. Consequently, if we robustly remove outliers and group inliers belonging to the same plane, we can compute the parameters directly from the set of inliers. Tensor voting achieves this in this paper.

From each correspondence (x,y) and (x',y'), we define two 4D spaces by decoupling the correspondence as $(x,y,x',1)$ and $(x,y,y',1)$. Each 4D space is orthogonal and isotropic, and each 4D point $(x,y,x',1)$ or $(x,y,y',1)$ lies on a plane parameterized by $(a,b,-1,t_x)$ and $(c,d,-1,t_y)$ respectively.

3.2. Outlier Removal
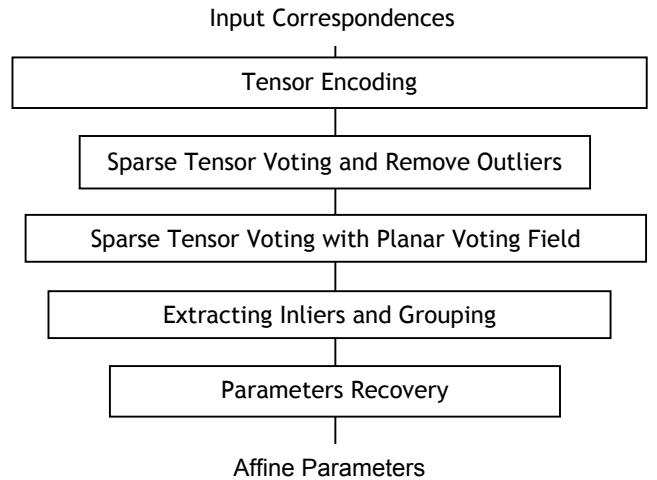
Input Correspondences

Figure 1. Recovering affine parameters.

Figure 1 shows the flowchart for recovering affine parameters from a set of correspondences. At first, input correspondences are converted to points in a decoupled joint

image space and encoded into a 4D ball tensor defined by the following eigenvalues and eigenvectors:

$$\lambda_1=\lambda_2=\lambda_3=\lambda_4=1$$
$$e_1=(1,0,0,0)\ e_2=(0,1,0,0)\ e_3=(0,0,1,0)\ e_4=(0,0,0,1)$$

During the first sparse voting, each point collects votes from its neighbours and characterizes a principal direction defining a 2D plane. The normal orientation of 2D plane is defined by eigenvector e_1 associated to the largest eigenvalue of the decomposed tensor. The saliency of the extracted plane is given by $(\lambda_1 - \lambda_2)$ and characterizes the support of the neighbours to the plane. Therefore, isolated random noise has small saliency due to little support from neighbours. During the second voting, voting is performed with a planar voting field with a narrow angle and wider neighbourhood derived from the obtained normal orientation. At this step, only highly salient points (defined by the median threshold of the saliency values from the first voting) participate in the voting. The planar voting field allows to enforce a global plane constraint. After the second voting, outlier rejection is performed based on mean value computed from salient points. This outlier removal step is followed by a clustering of inliers that lie on the same plane. The grouping starts with the most salient point and cluster points having the same normal direction and similar distance of *x-x'* or *y-y'*. If un-clustered inliers remain, the grouping step iterates using the next highest saliency. This allows clustering multiple planes in the 4D space corresponding to different affine motion.

3.3. Parameter Recovery

For each set of clustered inliers, we estimate corresponding affine motion parameters. Let the selected inliers in two joint-image space represented by $q_{xi} = (x_i, y_i, x_i', 1)$ for estimating parameters *a, b, t_x* and $q_{yi} = (x_i, y_i, y_i', 1)$ for *c, d, t_y* where *i* is the index for each inlier and *n* is the number of inliers in the cluster. Then, we compose a stacked matrix of inliers m_x and m_y, as $m_x{}^T = \begin{pmatrix} q_{x1}^T & q_{x2}^T & \cdots & q_{xn}^T \end{pmatrix}$ and $m_y{}^T = \begin{pmatrix} q_{yi}^T \end{pmatrix}_{i=1..n}$

For estimating parameters *a, b, t_x*, let $M_x = m_x{}^T \cdot m_x$ then

$$M_x \cdot p_x = \begin{pmatrix} \sum_n x^2 & \sum_n xy & \sum_n xx' & \sum_n x \\ \sum_n xy & \sum_n y^2 & \sum_n yx' & \sum_n y \\ \sum_n xx' & \sum_n yx' & \sum_n x'^2 & \sum_n x' \\ \sum_n x & \sum_n y & \sum_n x' & n \end{pmatrix} \begin{pmatrix} a \\ b \\ -1 \\ t_x \end{pmatrix} = 0$$

where M_x is the covariance matrix of the inliers and p_x is the parameters representation in the corresponding joint-image subspaces.

The parameters of the affine transform are therefore characterized by the eigenvector associated to the smallest eigenvalue of the covariance matrix. Conceptually this

method acts like conventional parameter estimation. However, we showed that encoding the metric of the joint-image spaces in the tensor voting formalism allows us to perform outliers removal and multiple affine motion estimation. In this paper, we only show a case for estimating parameters *a, b, t_x*. Parameters *c, d, t_y* can be derived in the same way.

4. Experimental Results

Affine parameter estimation has been studied by a large number of authors and therefore the presentation of a new approach has to be presented by processing challenging situations and naturally compare it to the state-of-the art. We chose to compare our method the results obtained by RANSAC. However, in the first example we chose a synthetic example that cannot be processed by a RANSAC technique: multiple affine motions. We generated a set of correspondences from two different motions (black and red lines) and added random correspondences (green lines) with similar amplitude motion as illustrated in Figure 2. The ratio of correct to noisy (ie random) is 0.5. Blue x marked points are selected as inliers by the system. Grouping these inliers into different motion groups is done as describe in section 3.2.
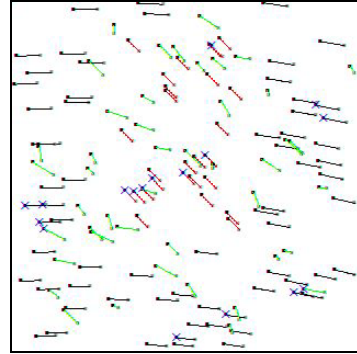

Figure 2. Synthetically generated correspondences.

Table 1 shows the comparison between the real affine parameter values and the parameters estimated by the proposed method.

	Affine parameters for black lines					
	A	*B*	*C*	*D*	*t_x*	*t_y*
Real	0.99	-.017	0.017	0.99	20.00	2.00
Estimated	1.00	-0.01	0.017	1.00	21.00	1.15
	Affine parameters for red lines					
	A	*B*	*C*	*D*	*t_x*	*t_y*
Real	1.00	0.00	0.00	1.00	10.00	10.00
Estimated	1.00	0.00	0.00	1.00	9.98	10.10

Table 1. Comparison estimated parameters to given real parameters.

In Figure 3 and Figure 6, we show a pair of frames extracted from two video sequences with moving objects in the scene. The purpose here is to compare the proposed method and RANSAC algorithm. We start by selecting feature points using a Harris corner detector with a low

threshold allowing to consider strong and weak corners. In Figure 4 and Figure 7 we show the correlation based initial matching. In Figure 5 and Figure 8 we show the image difference after compensating for the motion we have estimated using the proposed method. Here again we compare the results with the one obtained with the RANSAC method. One can clearly see a better image compensation resulting from the proposed technique. Especially in both Figure 5 in Figure 8, the background is more accurately registered.

5. Conclusion

In this paper, we proposed to use tensor voting to remove outliers within correspondences and robustly estimate affine parameters. Our approach contributed to define a 4D decoupled joint space from feature correspondences and showed that a 2D plane is a structure that affine motion constrains in the defined spaces. Also, we showed that the proposed method allows the computation of multiple affine motions simultaneously. Future work will investigate other parametric motions and global registration using a similar formalism.

Figure 3. Inputs (Walking Scene).

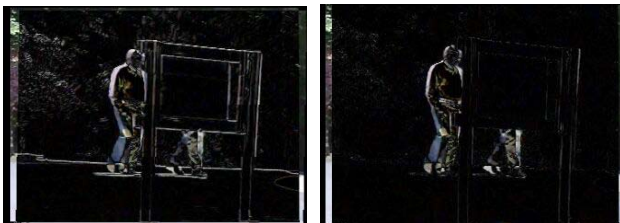Figure 4. Correlation-based initial correspondences.

Figure 5. Image difference after motion compensation by RANSAC(left) and our method(right).

Figure 6. Inputs (Basketball Scene).
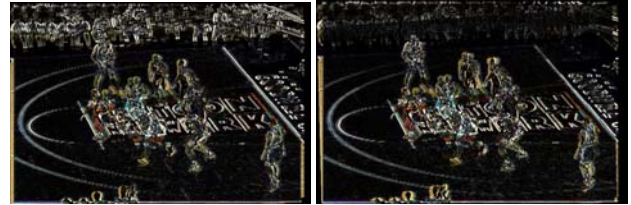
Figure 7. Correlation-based initial correspondences.

Figure 8. Image pixel differences after motion compensation by RANSAC(left) and our method(right).

References

[1] C. Tang, G. Medioni and M. Lee, Epipolar Geometry Estimation by Tensor Voting in 8D, ICCV 2000.

[2] G. Medioni, M. Lee and C. Tang, A Computational Framework for Feature Extraction and Segmentation, Elseviers Sci., Amsderstam, 2000.

[3] M. Black and P. Anandan, The robust estimation of multiple motions: Affine and piecewise-smooth flow fields, Tech. Report TR, Xerox PARC, Dec. 1993.

[4] M. Irani et al., Efficient Representations of Video Sequences and Their Applications, Signal Processing, *Vol. 8, No. 4, May 1996.*

[5] W. Tong, C. Tang and G. Medioni, Epipolar Geometry Estimation for Non-Static Scenes by 4d Tensor Voting, CVPR, 2001.

[6] S. Ayer, P. Schroeter and J. Bigun, Segmentation of moving objects by robust motion parameter estimation over multiple frames. ECCV, May, 1994.

[7] Z. Zhang, Determining the epipolar geometry and its uncertainty:A review. IJCV vol.27, no.2, pp.161-195, 1998.

[8] E. Kang, I. Cohen and G. Medioni, "A Graph-Based Global Registration for 2D Mosaics", ICPR 2000.

[9] A. Lacey, N. Pinitkarn and N. Thacker, An Evaluation of the Performance of RANSAC Algorithms for Stereo Camera Calibration, BMVC 2000.