The Impossibility of Affine Reconstruction from Perspective Image Pairs Obtained by a Translating Camera with Varying Parameters

Z. Y. Hu and F. C. Wu
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
P.O. Box 2728, Beijing 100080, P.R. China
{huzy,wufc}@nlpr.ia.ac.cn

Abstract

Contrary to the claim in the literature that the affine reconstruction is possible from two images captured by a translating camera with unknown and varying parameters, we show that such a reconstruction is in fact impossible. In other words, the knowledge that " the camera's motion between the two images is just a pure translation" is an insufficient piece of information for affine reconstruction.

1. Introduction

Affine reconstruction plays a very important role in camera self-calibration and the stratified 3D reconstruction [1]. It s well known that once the affine reconstruction is done, the camera calibration becomes a linear one (with fixed parameters)[2]. In the stratified 3D reconstruction, the affine reconstruction is the most delicate step compared with the projection reconstruction and the metric reconstruction [3]. In [4], it was shown that if the camera is only a translating one, and if the camera's intrinsic parameters do not change, then the affine reconstruction can be done with 5 corresponding image points between two images. In the case where the camera's intrinsic parameters do change, the authors claimed that " which can be dealt with, but complicate matters". However, in this short note, we show that it is in fact impossible to obtain an affine reconstruction from a pair of images captured by a translating camera with varying parameters. Here by " varying parameters", we mean all the five parameters under the pinhole model are subject to change.

Before elaborating on the issue, we first recall that given two images $I, I^{'}$, denoting the epipole in $I^{'}$ as $\boldsymbol{e}^{'}$, then the affine reconstruction can be expressed as:

$$\begin{cases} \mathbf{P}_{1\mathbf{A}} \approx \begin{pmatrix} \mathbf{I} & \mathbf{0} \end{pmatrix} \\ \mathbf{P}_{2\mathbf{A}} \approx \begin{pmatrix} \mathbf{H} & \mathbf{e}^{'} \end{pmatrix} \end{cases} \tag{1}$$

where $\boldsymbol{H}$ is the homography of the plane at infinity (called hereinafter the infinite homography) , " $\approx$ " means the equality up to a scale, $\boldsymbol{I}$ is the identity matrix. This infinite homography is unknown but unique. In other words, the affine reconstruction is to determine this infinite homography. If this infinite homography can be uniquely determined (in the sense of up to a scale), then the affine reconstruction is possible. Otherwise, the affine reconstruction is impossible.

2. What can be drawn from a pure translation of camera motion

We have the following proposition:

The infinite homography $\boldsymbol{H}$ is an upper triangular matrix with all the 3 diagonal elements being positive if and only if the camera motion between the two images is a pure translation.

Proof: Assume the camera matrices with the first and second images (captured before and after the translation) are respectively

$$\boldsymbol{K} = \begin{pmatrix} f_u & s & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{K}^{'} = \begin{pmatrix} f_u^{'} & s^{'} & u_0^{'} \\ 0 & f_v^{'} & v_0^{'} \\ 0 & 0 & 1 \end{pmatrix}$$

with $f_u > 0, f_v > 0, f_u^{'} > 0, f_v^{'} > 0$

(if part) If the camera motion between the two images is a pure translation $\boldsymbol{t}$, then the two projection matrices in the Euclidean space should be:

$$\begin{cases} \boldsymbol{P_{1E}} \approx \boldsymbol{K}(\boldsymbol{I} \quad \boldsymbol{0}) \\ \boldsymbol{P_{2E}} \approx \boldsymbol{K}^{'}(\boldsymbol{I} \quad \boldsymbol{t}) \end{cases} \tag{2}$$

then the infinite homography is:

$$\boldsymbol{H} \approx \boldsymbol{K}^{'}\boldsymbol{K}^{-1}$$

$$= \begin{pmatrix} \frac{f_u^{'}}{f_u} & \frac{s^{'}f_u - sf_u^{'}}{f_u f_v} & u_0^{'} - \frac{v_0 s^{'}}{f_v} + \frac{f_u^{'}(sv_0 - u_0 f_v)}{f_u f_v} \\ 0 & \frac{f_v^{'}}{f_v} & v_0^{'} - \frac{f_v^{'}}{f_v} v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

Hence $\boldsymbol{H}$ is an upper-triangular matrix and all the 3 diagonal elements are positive.

(only if part). If matrix $\boldsymbol{H}$ is an upper-triangular matrix with all the 3 diagonal elements being positive, since the general form of the infinite homography is $\boldsymbol{K}^{'}\boldsymbol{R}\boldsymbol{K}^{-1}$, where $\boldsymbol{R}$ is an unknown 3D rotation matrix, then from $\boldsymbol{H} = \lambda\boldsymbol{K}^{'}\boldsymbol{R}\boldsymbol{K}^{-1}$, we know $\lambda\boldsymbol{R} = (\boldsymbol{K}^{'})^{-1}\boldsymbol{H}\boldsymbol{K}$ must be an upper-triangular matrix also, and all the 3 diagonal elements must also be positive. It can be easily verified by construction that a rotation matrix, being an upper-triangular matrix and all its 3 diagonal elements being positive, must be an identity matrix, i.e., $\boldsymbol{R} = \boldsymbol{I}$. In other words, in this case, the camera's motion must be a pure translation, without any rotation involved.

3. The impossibility of affine reconstruction

Based on the proposition in Section 2, given two images captured by a translating camera with varying parameters, if we can find at least two upper-triangular matrices $\boldsymbol{H}, \boldsymbol{H}^{'}$ with all their diagonal elements being positive such that $[\boldsymbol{e}^{'}]_{\times}\boldsymbol{H} \approx [\boldsymbol{e}^{'}]_{\times}\boldsymbol{H}^{'} \approx \boldsymbol{F}$, where $\boldsymbol{F}$ is the fundamental matrix between the two given images and $[\boldsymbol{e}^{'}]_{\times}$ is the antisymmetric matrix defined by the epipole $\boldsymbol{e}^{'}$, then the following two reconstructions:

$$\mathrm{A}_1: \begin{cases} \boldsymbol{P_1} \approx (\boldsymbol{I} \quad \boldsymbol{0}) \\ \boldsymbol{P_2} \approx (\boldsymbol{H} \quad \boldsymbol{e}^{'}) \end{cases}$$

$$\mathrm{A}_2: \begin{cases} \boldsymbol{P}_1^{'} \approx (\boldsymbol{I} \quad \boldsymbol{0}) \\ \boldsymbol{P}_2^{'} \approx (\boldsymbol{H}^{'} \quad \boldsymbol{e}^{'}) \end{cases}$$

both encapsulate all available information (a translating camera and image correspondences). Since the affine reconstruction under the form (1) must be unique, it means that in this case, the affine reconstruction is impossible.

Based on the above reasoning, the problem is reduced to find out all possible upper-triangular matrices $\boldsymbol{H}$ with positive diagonal elements subject to:

$$[\boldsymbol{e}^{'}]\boldsymbol{H} \approx \boldsymbol{F} \tag{4}$$

Clearly, $\boldsymbol{H}$ is not unique in (4). This is because if $\boldsymbol{H}$ is a solution, $\boldsymbol{H}^{'} = \boldsymbol{H} + \boldsymbol{e}^{'}\boldsymbol{x}^{\boldsymbol{T}}$ must also be a solution, where $\boldsymbol{x}$ is an arbitrary 3-vector.

Since additionally if $\boldsymbol{H}$ is an upper-triangular matrix, $\boldsymbol{H} + \boldsymbol{e}^{'}\boldsymbol{x}^{\boldsymbol{T}}$ must also be an upper-triangular matrix, some additional constraints can be enforced on vector $\boldsymbol{x}$. There are two possible cases:

Case 1: If $\boldsymbol{e}^{'}$ is not at infinity, then $(\boldsymbol{e}^{'})^T = (e_1 \quad e_2 \quad 1)$, then from

$$\mathbf{H} + \mathbf{e}^{'}\mathbf{x} = \begin{pmatrix} h_1 & h_2 & h_3 \\ 0 & h_4 & h_5 \\ 0 & 0 & h_6 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ 1 \end{pmatrix} (x_1 \quad x_2 \quad x_3)$$

$$\approx \begin{pmatrix} h_1^{'} & h_2^{'} & h_3^{'} \\ 0 & h_4^{'} & h_5^{'} \\ 0 & 0 & h_6^{'} \end{pmatrix} = \mathbf{H}^{'}$$

we have $x_1 = x_2 = 0$, and x_3 can be an arbitrary positive number. In other words, the infinite homography cannot be uniquely determined in this case, it is a function of one free parameter.

Case 2: If $\boldsymbol{e}^{'}$ is at infinity, $(\boldsymbol{e}^{'})^T = (e_1 \quad e_2 \quad 0)$, then from

$$\mathbf{H} + \mathbf{e}^{'}\mathbf{x} = \begin{pmatrix} h_1 & h_2 & h_3 \\ 0 & h_4 & h_5 \\ 0 & 0 & h_6 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ 0 \end{pmatrix} (x_1 \quad x_2 \quad x_3)$$

$$\approx \begin{pmatrix} h_1^{'} & h_2^{'} & h_3^{'} \\ 0 & h_4^{'} & h_5^{'} \\ 0 & 0 & h_6^{'} \end{pmatrix} = \mathbf{H}^{'}$$

we have $x_1 = 0$, and x_2, x_3 are two free parameters. In this case, the infinite homography is a function of two free parameters.

Combining Case 1 and Case 2, we know that by merely knowing that " camera is undergoing a pure translation" is insufficient for an affine reconstruction from a pair of perspective images.

Before ending this note, we would make the following two remarks:

Remark 1: As we have proved, if all the five intrinsic parameters of the camera are subject to change, then an affine reconstruction is impossible. However, if some of the parameters are known, for example, the principle points $(u_0 \quad v_0), (u_0^{'} \quad v_0^{'})$ are known, then an affine reconstruction becomes possible.

Remark 2: In [4], it was shown that if the camera matrix is fixed and if 5 point correspondences are known, then an affine reconstruction can be obtained. In fact, in this case, if only two point correspondences are available, an affine reconstruction can be done. This is because if the camera is fixed, the infinite homography $\boldsymbol{H}$ in (1) must be the identity matrix. In addition, by two point correspondences, the epipole $\boldsymbol{e}^{'}$ can be determined by intersecting the two lines determined by the two pairs of image corresponding points.

Acknowledgments

This work was supported by "973" Program (G1998030502), The National Science Foundation of China (60033010, 69975021).

References

[1]: R.Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press 2000.

[2]: R.Hartley, Self-Calibration of Stationary Cameras, International Journal of Computer Vision 22(1), pp.5-23, 1997.

[3]: R. Hartley, E. Hayman, L. de Agapito, and I. Reid, Camera Calibration and the Search for Infinity, International Conference on Computer Vision, Bombay, India, pp.510-516, 1999.

[4]: T. Moons, L Van Gool, M. Proesmans, E. Pauwel, Affine Reconstruction from Perspective Image Pairs Obtained by Translating Camera, Application of Invariance in Computer Vision, LNCS825, Springer-Verlag, 1994.