A Note on Polynomials Based Image Registration[1)]

HUANG Feng-Rong HU Zhan-Yi

(*National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing* 100080)

(E-mail: frhuang@nlpr.ia.ac.cn, huzy@nlpr.ia.ac.cn)

Abstract It is shown that the polynomials based image registration, which is widely used in remote sensing field, does not have a sound mathematical basis. In fact, there seems no theoretical basis for the polynomials based transform to outperform the affine transformation, a much simpler one, in image registration. If the transformation functions are polynomials of order n, the corresponding scene is shown to be in general the intersection of two curved surfaces of order $n+1$, in other words, a space curve. In some special cases, the scene is approaching to a plane. To our knowledge, such results did not appear in the literature previously.

Key words Image registration, polynomial transformation, affine transformation, homography

1 Introduction

Image registration is a fundamental task in computer vision and remote sensing[1~4] to match two or more images taken at different time, or from different sensors, or from different viewpoints. Image registration is essentially to determine the correspondences of image points across images. The general approach to image registration is usually carried out in two steps. At first, a number of prominent points are selected and their correspondences are established manually or by some automatic means. These corresponding points are called control points and used later to determine the transformation functions to match the rest points in images. Over years, in order to offset the inaccuracy of the affine transformation based matching, people, in particular, the people in remote sensing field, popularly used polynomials as the transformation functions, a natural extension of the affine transformation, to match image points. Hence problems come. Does the polynomial transformation provide a viable alternative for the affine one? Or is the rationale of the polynomial transformation sound enough in image registration? The objective of this note is to clarify these matters. Our results from a theoretical analysis show that in general, the polynomial transformation is not a viable candidate in image registration. In fact, even if the space points lie on a plane, the corresponding image points cannot be related by a polynomial transformation. Besides, the polynomial transformation will not necessarily outperform the affine transform in image registration.

The paper is organized as follows. In Section 2, some preliminaries are given. Main results will be elaborated in Section 3. Section 4 is a study on the transformation with a planar scene. Some concluding remarks are listed at the end of the paper.

2 Preliminaries

2.1 Notation

The camera model employed here is of the pinhole one. The following notation is used in this paper: an image point is denoted by $\boldsymbol{u} = [u, v]^{\mathrm{T}}$, and its homogeneous coordinates are denoted by $\tilde{\boldsymbol{u}} = [u, v, 1]^{\mathrm{T}}$ and a 3D point is denoted by $\boldsymbol{X} = [X_w, Y_w, Z_w]^{\mathrm{T}}$. Then the imaging process from a 3D point X to its 2D image u can be expressed as

$$\lambda\tilde{\boldsymbol{u}} = K[P \quad \boldsymbol{t}]\begin{bmatrix}\boldsymbol{X}\\1\end{bmatrix} \tag{1}$$

where λ is a non-zero scale factor, $(R \quad \boldsymbol{t})$ are the rotation matrix and translation vector from the world system to the camera system, and K is the camera intrinsic matrix. If the camera intrinsic matrix is

1) Supported by National Natural Science Foundation of P. R. China (60175009, 60121302). Corresponding author: Hu Zhan-Yi

Received May 19, 2003; in revised form November 3, 2003

known, then

$$\lambda\tilde{\boldsymbol{u}}_0 = \lambda K^{-1}\tilde{\boldsymbol{u}} = [R \quad \boldsymbol{t}]\begin{bmatrix}\boldsymbol{X}\\1\end{bmatrix} \tag{2}$$

$\tilde{\boldsymbol{u}}$ is called the normalized coordinates. For the sake of clarity, image points will be assumed in the normalized coordinates throughout this paper. Equivalently, the camera matrix K will be assumed to be the identity matrix.

2.2 Homography from a space plane to its image

A space plane, without loss of generality, can be assumed to be $Z_w = 0$. By denoting the i^{th} column of the rotation matrix R by $\boldsymbol{r}_i$, t hen from (1), we have

$$\lambda\tilde{\boldsymbol{u}} = [\boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \boldsymbol{r}_3 \quad \boldsymbol{t}]\begin{bmatrix}X_w\\Y_w\\0\\1\end{bmatrix} = [\boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \boldsymbol{t}]\begin{bmatrix}X_w\\Y_w\\1\end{bmatrix}$$

Therefore, a plane point $\boldsymbol{X}$ and its image $\boldsymbol{u}$ are related by a plane homography H:

$$\lambda\tilde{\boldsymbol{u}} = H\begin{bmatrix}X_w\\Y_w\\1\end{bmatrix} \qquad H = [\boldsymbol{r}_1 \quad \boldsymbol{r}_2 \quad \boldsymbol{t}] = \begin{bmatrix}h_{11} & h_{12} & h_{13}\\h_{21} & h_{22} & h_{23}\\h_{31} & h_{32} & h_{33}\end{bmatrix} \tag{3}$$

The 3×3 matrix H is only defined up to a scale factor.

3 Scene structure from polynomial transformation

3.1 Scene structure from affine transformation

Suppose $\{(u_i \quad v_i) \leftrightarrow (u_i' \quad v_i')\ i = 1, 2, \cdots, N\}$ is a set of corresponding image points. By " affine transformation", we mean that the corresponding pairs of image points satisfy an affine transformation, such as

$$\begin{pmatrix}u_i'\\v_i'\end{pmatrix} = \begin{pmatrix}b_1\\b_2\end{pmatrix} + \begin{pmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\end{pmatrix}\begin{pmatrix}u_i\\v_i\end{pmatrix} \qquad i = 1, 2, \cdots, N \tag{4}$$

Many cases show that the affine transformation based registration brings gross errors. In Section 3.2, it will be shown that only those space points with special structure can produce image points satisfying an affine transformation.

3.2 Scene structure from polynomial transformation

Clearly, a natural extension of the affine transformation is a polynomial one such as

$$\begin{cases} u' = \displaystyle\sum_{i=0}^{n}\sum_{j=0}^{n-i} a_{ij}u^i v^j \\ v' = \displaystyle\sum_{i=0}^{n}\sum_{j=0}^{n-i} b_{ij}u^i v^j \end{cases} \tag{5}$$

The question is that whether such an extension can indeed alleviate the matching problem. Before answering such a question, let us at first look at what the corresponding scene should behave if the corresponding images do satisfy a polynomial transformation of order n. We have the following proposition.

Proposition 1. The corresponding scene is the intersection of two curved surfaces of order $n+1$ if the transformation functions in (5) are polynomials of order n.

Proof. For the simplicity purpose, the two camera systems are additionally assumed to be only under a translation $\boldsymbol{t} = [t_1, t_2, t_3]^{\text{T}}$, *i.e.*, no rotation is involved. Hence the imaging process from a 3D point $\boldsymbol{X}$ in space to its image $\tilde{\boldsymbol{u}} = [u, v, 1]^{\text{T}'}$ in the first image and $\tilde{\boldsymbol{u}}' = [u', v', 1]^{\text{T}}$ in the second image can be expressed as

$$\lambda\tilde{\boldsymbol{u}} = [I \quad 0]\begin{bmatrix}\boldsymbol{X}\\1\end{bmatrix} = \boldsymbol{X} \tag{6}$$

$$\lambda'\tilde{\boldsymbol{u}}' = [I \quad \boldsymbol{t}]\begin{bmatrix}\boldsymbol{X}\\1\end{bmatrix} = \boldsymbol{X} + \boldsymbol{t} \tag{7}$$

From (6),

$$u = \frac{X_w}{Z_w}, \quad v = \frac{Y_w}{Z_w} \tag{8}$$

From (7),

$$u' = \frac{X_w + t_1}{Z_w + t_3}, \quad v' = Y_w + t_2 Z_w + t_3 \tag{9}$$

Then substituting (8) and (9) into (5), we have

$$\frac{X_w + t_1}{Z_w + t_3} = \frac{1}{Z_w^n} \sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{ij} X_w^i Y_w^j Z_w^{n-i-j}, \quad \frac{Y_w + t_1}{Z_w + t_3} = \frac{1}{Z_w^n} \sum_{i=0}^{n} \sum_{j=0}^{n-i} b_{ij} X_w^i Y_w^j Z_w^{n-i-j}$$

or

$$(Z_w + t_3)\left(\sum_{i=0}^{n} \sum_{j=0}^{n-i} a_{ij} X_w^i Y_w^j Z_w^{n-i-j} \right) - Z_w^n (X_w + t_1) = 0$$

$$(Z_w + t_3)\left(\sum_{i=0}^{n} \sum_{j=0}^{n-i} b_{ij} X_w^i Y_w^i Z_w^{n-i-j} \right) - Z_w^n (Y_w + t_2) = 0$$

The above equations indicate that the corresponding surfaces are in general two complicated curved surfaces of order $(n+1)$, *i.e.*, the scene is the intersection of these two curved surfaces of order $(n+1)$, *i.e.*, generally a space curve.

If the order is equal to 2, *i.e.*, $n = 2$, then

$$u' = a_1 u_2 + a_2 v^2 + 2a_3 uv + 2a_4 u + 2a_5 v + a_6 = \tilde{\boldsymbol{u}} \begin{bmatrix} a_1 & a_3 & a_4 \\ a_3 & a_2 & a_5 \\ a_4 & a_5 & a_6 \end{bmatrix} \tilde{\boldsymbol{u}} = \tilde{\boldsymbol{u}}^{\mathrm{T}} A \tilde{\boldsymbol{u}} \tag{10}$$

$$v' = b_1 u^2 + b_2 v^2 + 2b_3 uv + 2b_4 u + 2b_5 v + b_6 = \tilde{\boldsymbol{u}} \begin{bmatrix} b_1 & b_3 & b_4 \\ b_3 & b_2 & b_5 \\ b_4 & b_5 & b_6 \end{bmatrix} \tilde{\boldsymbol{u}} = \tilde{\boldsymbol{u}}^{\mathrm{T}} B \tilde{\boldsymbol{u}} \tag{11}$$

Similarly as shown in the above, we have

$$(Z_w + t_3) \times \boldsymbol{X}^{\mathrm{T}} A \boldsymbol{X} - Z_w^2 \times (X_w + t_1) = 0 \tag{12}$$

$$(Z_w + t_3) \times \boldsymbol{X}^{\mathrm{T}} B \boldsymbol{X} - Z_w^2 \times (Y_w + t_2) = 0 \tag{13}$$

(12) and (13) are two curved surfaces of order 3 in space, then the scene is the intersection of the two curved surfaces of order 3.

Please notice that the above proof of Proposition 1 can be reversed. As shown in Fig. 1, we denote the intersection of the two curved surfaces as P, which is usually a space curve, and denote the images of P on the two image planes as p in the first image and p' in the second image, which are two plane curves, respectively. Then any point on p in the first image and its corresponding point on p' satisfy the polynomial transformation functions in (5). As a result, in practice, there exist only a few image points in the two images that can satisfy the polynomial transformation.

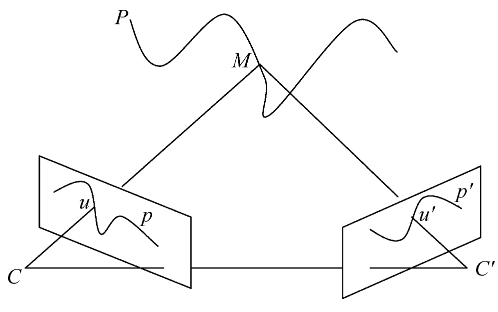


Fig. 1 Only a few image points can satisfy the polynomial transformation

From Proposition 1, we know that only space points lie on the intersection of some complicated curved surfaces, *i.e.*, a complicated space curve, the resulting images satisfy polynomial transformations. In general, the corresponding points cannot be formulated as polynomial transformations. Such results are hardly surprising. The rationale behind any polynomial transformations is that the image points from two images should have a one-to-one mapping defined by a function, no matter how complicated such a function is. As we know, polynomial functions are merely the approximations of the Taylor expansion series of the function. However, from the epipolar geometry which has been extensively investigated in recent years in computer vision, we know that such a one-to-one mapping does not exist in general. Without any a priori knowledge on the scene and camera motion, what we can at best know is the fact that the corresponding point must lie on a line (called epipolar line), *i.e.*, the problem of point correspondence is in essence the one of one-to-many mapping. As a result, we here claim that the polynomials based image registration is lack of sound mathematical basis. Perhaps, in some special cases, it works well. But any blithe generalization is doomed to failure.

4 A special case of study

Proposition 1 says that in the general case the resulting scene is the intersection of two curved surfaces of order $(n+1)$. In this section a special case, *i.e.*, a planar surface, is considered. As shown in Appendix, the images points projected from a space plane are related by a homography. In this case, a one to one mapping does exist. More explicitly, suppose $\{\tilde{\boldsymbol{u}} = (u_i \quad v_i \quad 1)^{\mathrm{T}} \leftrightarrow \tilde{\boldsymbol{u}}'_i = (u'_i \quad v'_i \quad 1)\ i = 1, 2, \cdots, N\}$ is a set of corresponding image points projected from a space plane. Then

$$\lambda\tilde{\boldsymbol{u}}' = \lambda\begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = H\tilde{\boldsymbol{u}} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix}\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{14}$$

i.e.,

$$u' = \frac{h_{11}u + h_{12}v + h_{13}}{h_{31}u + h_{32}v + h_{33}} \tag{15}$$

$$v' = \frac{h_{21}u + h_{22}v + h_{23}}{h_{31}u + h_{32}v + h_{33}} \tag{16}$$

If $h_{31}u + h_{32}v + h_{33} \neq 0$, then (15) and (16) can be expanded at (u_0, v_0) into polynomials of infinitely large order as

$$\begin{cases} u' = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} a_{ij}x^i y_j \\ v' = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} b_{ij}x^i y_j \end{cases} \tag{17}$$

where $x = u - u_0, y = v - v_0$. (17) says that the image points from a space plane satisfy polynomial transformation between a pair of two images, the order of the polynomials is infinitely large.

As shown in Fig. 2, the geometric interpretation of the condition $h_{31}u + h_{32}v + h_{33} \neq 0$ is that the intersecting line between the space plane and the focal plane of the second camera (*i.e.*, the plane going through the optical center of the second camera and parallel to its image plane) should lie outside the viewing cone of the first camera. In other words, the projection of this intersecting line should lie outside the first image. This is because $h_{31}u + h_{32}v + h_{33} = 0$ is an image line in the first image. By (14), this image line is transformed to the line at infinity in the second image and the line at infinity in the second image must be generated by the focal plane of the second camera. In other word, the corresponding line in space is the intersection line of the space plane and the focal plane of the second camera. As shown in Fig. 2, π_s is the space plane, plane π_f is the focal plane of the second camera. The space line L is the intersecting line between π_s and π_f. If line L does not lie

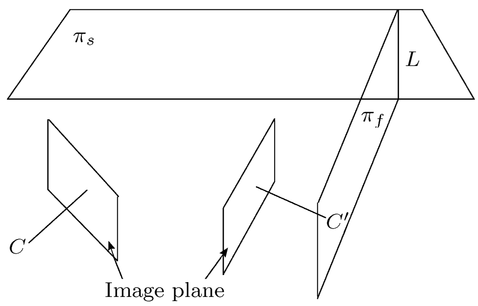
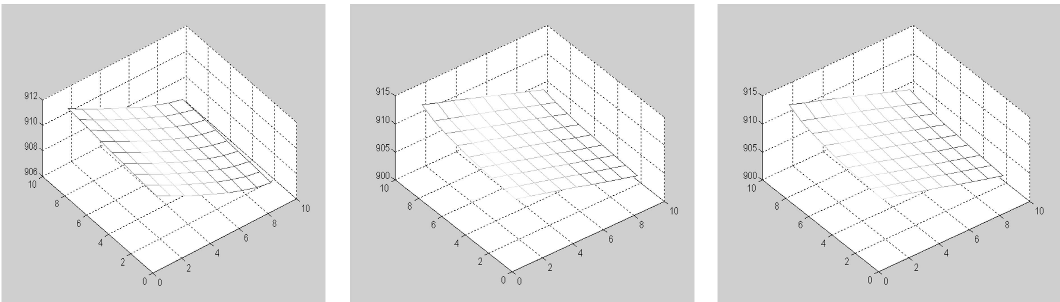


Fig. 2 In order for (15) (16) to be expansible the intersection line h should be outside of the view cone of the first camera

within the viewing cone of the first camera, the resulting homography can be expanded by the Taylor expansion.

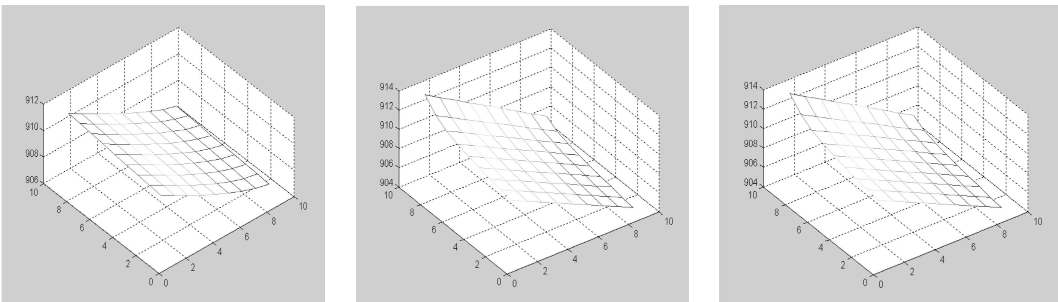
Computer simulations. Here two sets of simulations are reported. Two space planes, with one far away from the two cameras (the distance of the space plane to the first camera is about ten times of the translation magnitude between the two cameras), and the other close to the two cameras (the distance of the space plane to the first camera is about three times of the translation magnitude between the two cameras). The reason of selecting such planes is that for the first one, its projective effect is relatively small due to large distances to the two cameras. But for the second one, the projective distortion is significant. The details are as follows.

1) The Space plane is far away from the two cameras. The plane is $Z_w = 0$, *i.e.*, the XY plane of the world system. The setup of the first camera to the world system is: rotation axis is $\boldsymbol{r} = [20 \quad 45 \quad 100]^{\mathrm{T}}$, rotation angle $\pi/6$, and translation $\boldsymbol{T}_1 = [30 \quad -10 \quad 900]^{\mathrm{T}}$. The translation between the two cameras is $T_2 = [3 \quad 1 \quad 90]^{\mathrm{T}}$. We calculate the homography as shown in [1], then we expand the homography into polynomials of orders 10, 14 and 19 at point (0,0) and (10,6), respectively by Taylor expansion, *i.e.*, (15) and (16) are expanded into polynomials of orders 10, 14 and 19 at point (0, 0) and (10,6), respectively. Finally, in each case, we reconstruct 81 real space points of the scene using the method in Section 3.2. The results with the homography expanded at point (0,0) are shown in Fig. 3, and those at point (10,6) are shown in Fig. 4.



(a) (b) (c)

Fig. 3 (a) Intersection of curved surfved surfaces of order 11, (b) Intersection of curved surfaces of order 15, (c) Intersection of curved surfaces of order 20



(a) (b) (c)

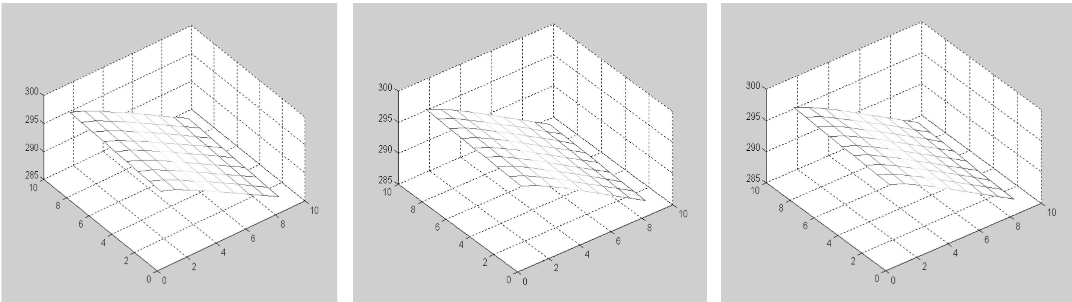
Fig. 4 (a) Intersection of curved surfved surfaces of order 11, (b) Intersection of curved surfaces of order 15, (c) Intersection of curved surfaces of order 20

Table 1 The average distance from the reconstructed points to the space plane is listed in

Far away space plane	order 11	order 15	order 20
Expand at point (0,0)	2.2575	1.4956	1.4956
Expand at point (10,6)	2.3083	1.4986	1.4986

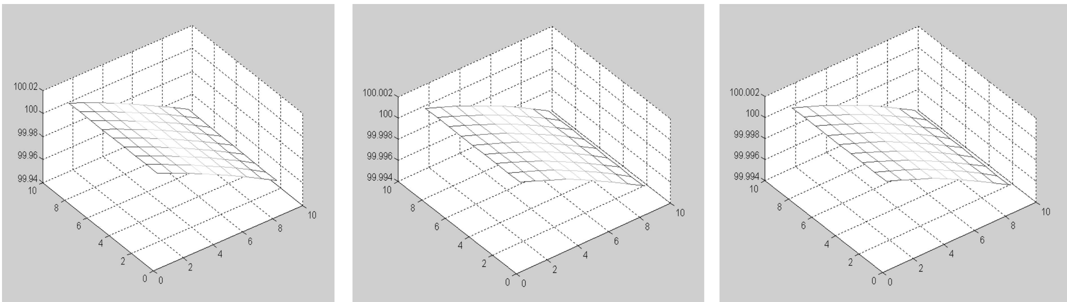
2) The space plane is close to the two cameras.

The plane is under the world system. The setup of the first camera to the world system is: rotation axis is $\boldsymbol{r} = [2 \quad 10 \quad 5]^{\mathrm{T}}$, rotation angle $\pi/6$, and translation $T_1 = [-15 \quad -10 \quad 300]^{\mathrm{T}}$. The translation between the two cameras is $T_2 = [5 \quad 4 \quad -100]^{\mathrm{T}}$. Once again, we calculate the homography with the method in [1], then we expand the homography into polynomials of orders 10,14 and 19 at point (0,0) and (10,-7) respectively. Finally, in each case, we reconstruct 81 real space points of the scene using the method in Section 3.2. The results with the homography expanded at point (0,0) are shown in Fig. 5 and those at point (10,-7) are shown in Fig. 6.



(a) (b) (c)

Fig. 5 (a) Intersection of curved surfved surfaces of order 11, (b) Intersection of curved surfaces of order 15, (c) Intersection of curved surfaces of order 20



(a) (b) (c)

Fig. 6 (a) Intersection of curved surfved surfaces of order 11, (b) Intersection of curved surfaces of order 15, (c) Intersection of curved surfaces of order 20

Table 2 The average distance from the reconstructed points to the space plane is listed

Near space plane	order 11	order 15	order 20
Expand at point (0,0)	2.8190	2.8089	2.8089
Expand at point (10,-7)	171.2705	170.2601	170.2601

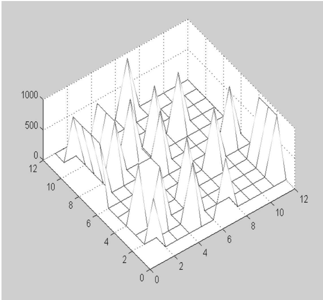
Remarks.

1) In some special cases, the scene is approaching to a plane, *i.e.*, the intersection of two curved surfaces of order n may approach to a plane. The simulations confirm this. From Fig. 2∼5, we can see that the reconstructed scene is almost a planar one.

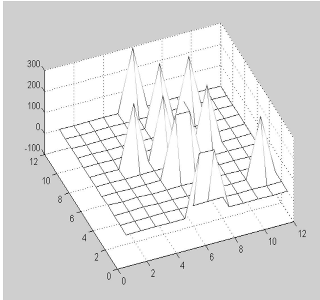
2) Although the homography can be expanded at u_0, for different image point $\boldsymbol{u}$, the accuracy of its corresponding reconstructed scene point is different since $|\boldsymbol{u} - \boldsymbol{u}_0|$ is different.

3) From Table 1 and Table 2, if a plane is close to the two cameras, its reconstructed error by polynomials transformation is larger than that of a faraway plane. This is consistent with the conventionalism since the closer a plane, the more severe the projective distortion.

4) For two planes we studied, we find that if we expand (15) and (16) into polynomials of lower order, say, less than 10, then the reconstructed points are quite sparse and the intersection of two curved surfaces is not approaching to a plane, as shown in Fig. 7. This shows that for planar scene, using polynomial functions of lower order to match images is not viable.



Space plane is far away



Space plane is close

Fig. 7 Intersection of curved surface of order 5 (expand at point (0, 0))

5 Conclusions

In this short note, we show that generally speaking, if the corresponding image points from two images are governed by a polynomial transformation of order n, the corresponding space points must lie on the intersection of two curved surfaces of order $(n + 1)$. In some special cases, the scene is approaching to a plane. Furthermore, we show that there seems no sound theoretical basis for the polynomials based image registration. In fact, even if the space points lie on a plane, the corresponding image points cannot be formulated by a polynomial transformation.

References

1 Brown L G. A survey of image registration techniques. *ACM computing surveys*, 1992, **24**(4): 325~376

2 Goshtasby A. Piecewise linear mapping functions for image registration. *Pattern Recognition*, 1987, **20**(5): 525~533

3 Goshtasby A. Image registration by local approximation methods. *Image and Vision Computing*, 1988, **6**(4): 255~261

4 Fogel D N. Image registration using multiquadric functions, the finite element method, Bivariate mapping polynomials and thin plate spline. Technical Report, National Center for Geographic Information and Analysis. Santa Barbara: University of California, 1996

HUANG Feng-Rong Ph. D. candidate at Institute of Automation, Chinese Academy of Sciences. She received her master degree from Heibei University of Technology in 2001. Her research interests include pattern recognition, camera calibration, 3D reconstruction, and image based measuring.

HU Zhan-Yi Professor at Institute of Automation, Chinese Academy of Sciences. He received his Ph. D. degree (Docteur d′Etat) in computer vision from University of Liege, Belgium, in 1993. His research interests include camera calibration and 3D reconstruction, active vision, geometric primitive extraction, vision guided robot navigation, and image based modeling and rendering.

Appendix

Relationship between two images of a space plane

Proposition. The relationship between two images of a space plane is a plane to plane homography.

Proof. From Section 2.2, we know that the map between a plane and its images is a plane to plane homography, and then, the map between the plane and its first image and its second image can be expressed as

$$\lambda\tilde{\boldsymbol{u}} = H_1 \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \qquad \lambda'\tilde{\boldsymbol{u}}' = H_2 \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix}$$

then,

$$s\tilde{\boldsymbol{u}}' = H_2 H_1^{-1}\tilde{\boldsymbol{u}} = H\tilde{\boldsymbol{u}}$$

where s is a non-zero scale factor. Therefore, the relationship between the two images of a space plane is also a plane to plane homography. □