Application of the Total Least Squares Procedure to Linear View Interpolation

[1]D M Kennedy, [2]B F Buxton and [1]J H Gilby
[1]Sira Limited, South Hill, Chislehurst, Kent BR7 5EH
[2]Department of Computer Science, University College London, Gower Street, London WC1E 6BT
dawn.kennedy@ptp.sira.co.uk, b.buxton@cs.ucl.ac.uk, john.gilby@sira.co.uk

Abstract

It is shown that, in comparison to the results obtained from a conventional least squares approach, a total least squares solution leads to significant improvements in the geometry and appearance of images synthesised in a linear combination of views procedure. Use of the total least squares criterion is appropriate when errors on the control points are independently and identically distributed between the basis images and the target image being synthesised. When this is not be the case it is pointed out that the generalised total least squares procedure should be used. A synthetic object is used to evaluate the improvement in geometric accuracy obtained by use of the total least squares solution in comparison to a classical least squares method. Simulated and real images of laboratory test objects are similarly used to illustrate the improvement in appearance of images reconstructed by means of the total least squares procedure.

1 Introduction

The problem of view synthesis, in which novel views of objects and scenes are constructed from a small set of basis views, has received much attention recently [3-11]. Usually, but not invariably, the basis views are existing images and the reconstruction is based on a form of image interpolation or morphing that originated with the work of Ullman and Basri on object recognition. They first showed [1] that it is possible, to a good approximation, to derive linear relationships between the geometry of three views of a scene. Since then the theory has been greatly extended, in particular by the application of geometric invariance principles [see for example 12 and 13] and development of the tri and quadrifocal tensors [14-16], but it appears that a subtle point, first noticed recently in computer vision in the context of estimating the fundamental matrix [17], the way image measurement errors are treated has not been addressed. The aim of this paper is to show how, for the approximate linear morphing relationships most frequently used in practice, a correct error analysis may be carried out and to evaluate the resulting improvements in the geometry and appearance of the reconstructed images.

BMVC 1999 doi:10.5244/C.13.31

2 Linear Combination of Views

Although the idea of a linear combination of views was originally proposed [1] for object recognition and derived under assumption of orthographic projection, it has since been realised [2] that the model may easily be derived under the assumption of affine or weak perspective projection. In this case, as described in [9]:

$$x = p_{11}X + p_{12}Y + p_{13}Z + p_{14}, \quad y = p_{21}X + p_{22}Y + p_{23}Z + p_{24} \tag{1}$$

where (x,y) are the image co-ordinates and (X,Y,Z) are the world co-ordinates.

Given two views with image co-ordinates (x',y') and (x'',y'') we have four equations of the form (1). Using three of the four equations we can derive expressions for X, Y, and Z that can then be substituted into equation (1). In the absence of errors in the image co-ordinates, a number of equivalent relationships may be obtained. For example, if we choose to retain x', y' and x'' this gives the equations first obtained by Ullman and Basri [1] by consideration of the changes in appearance of an object under orthographic projection:

$$x = a_1x' + a_2y' + a_3x'' + a_4, \quad y = b_1x' + b_2y' + b_3x'' + b_4 \tag{2}$$

The simplest way to use equations (2) is to regard them as a kind of image warping. Thus, given a set of four corresponding control points on an object in each of the three images it is possible to solve for the coefficients a_i, b_i and to regard the resulting transformation as a mapping from which, given two basis images or views $I'(x',y')$ and $I''(x'',y'')$, we can reconstruct the co-ordinates (x,y) in the third, target image $I(x,y)$.

In practice, however, the corresponding control points are likely to contain measurement errors, so the above procedure is not correct and likely to lead to very large errors. It would be better to use a least squares solution of the original equations (1) and use criteria such as the accuracy or conditioning of the solutions to select the most appropriate relationship of form (2). However, it has been shown [9,10] that, it is possible to use an over-complete set of equations of the form

$$x = a_1x' + a_2y' + a_3x'' + a_4y'' + a_5, \qquad y = b_1x' + b_2y' + b_3x'' + b_4y'' + b_5, \tag{3}$$

which have the advantage of being symmetrical in the basis view co-ordinates [9] and have been found to be numerically quite stable in practice [9-11]. Thus, unlike (2), equations (3) do not produce large errors if the target view is close to one of the basis views, in this case the second view, $I''(x'',y'')$.

Given a set of control points we can then solve for the coefficients a_i, b_i in (3) in a classical least squares sense. However, this assumes [18,19] that all the measurement errors are contained within the target view co-ordinates (x,y) and that the basis view co-ordinates (x',y') and (x'',y'') are error free. In fact, this is never the case. The whole point of such view interpolation schemes [6,7] is that the basis views are obtained from actual images and that other, usually novel, views are reconstructed from them without having explicitly to construct a 3D model of the object. There will thus always be errors on the basis view control point co-ordinates (x',y') and (x'',y''), even if they are selected by hand [9-11].

Moreover, there will also be errors on the target view co-ordinates (x,y). In test experiments [9-11] where the target views are themselves also real images used for a quantitative evaluation of the view interpolation, the errors on (x,y) will be similar to those on (x',y') and (x'',y''). Under these conditions, a total least squares error criterion is appropriate [18,20]. This is the case analysed in detail in the remainder of this paper as the target image to be reconstructed can be used to evaluate the procedure quantitatively.

However, we point out that even when a *novel* view is constructed using such an interpolation scheme, there will still be errors on the control co-ordinates (x,y). For example, if (x,y) are inferred from some kind of view interpolation [6] from (x',y') and (x'',y''), errors from the basis views will propagate to (x,y). Similarly, if the original viewpoints (ie. the camera co-ordinates) for the basis views are known or inferred, and the viewpoint of the novel view inferred or interpolated there will still be error propagation to (x,y). Finally, even if the coefficients a_i, b_i are interpolated directly [21], a similar fitting problem will still be encountered in parameterising the basis views.

In general therefore, there may be errors on all terms in (3) except the last two containing the constants a_5 and b_5, and these errors may be correlated. The technique capable of dealing with all such problems is the *generalised total least squares method* [22], introduced recently in computer vision research to deal with similar problems in calculating the fundamental matrix [17]. Its application is similar to that of the total least squares method used here, except that it leads to a generalised eigenproblem [23] whereas the total least squares requires only a conventional eigenproblem or singular value decomposition. Since the total least squares method is the more familiar, simpler and, after a simple transformation, appropriate for the quantitative evaluation of the view interpolation, it will be used in what follows. Before describing how the total least squares approach was used, we finally note for completeness that a similar error analysis should, in principle, be applied to all applications of the essential and fundamental matrices and tri- and quadrifocal tensors [16] although, unlike the approximate view interpolation considered here, these problems are all non-linear, requiring either an approximate linearisation of the error weighting [24] and/or an iterative solution.

3 Application of the total least squares method

In the experiments reported here, all the points in all views are found using the same method and therefore likely to be affected by the same errors. For complete symmetry between all six image co-ordinates, it is therefore best to seek linear relationships of the form:

$$l_1 x + l_2 y + l_3 x' + l_4 y' + l_5 x'' + l_6 y'' + l_7 = 0 . \quad (4)$$

We can now solve (4) under the assumption that the errors are independently and identically distributed among all the co-ordinates. If all the terms were affected by errors and the errors are independently and identically distributed it is possible to solve using a total least squares (TLS) technique [18,19,20,22]. However, the particular case we have here has a term l_7 at the end which is the co-effiecent to be multiplied by the constant 1, which does not have any measurement error associated with it. In principle, we should therefore solve (4) as a generalised total least squares (GTLS) problem [22,17], which effectively [22] uses matrix transformations to eliminate such terms. In our case, however, we may proceed more simply by eliminating this term directly by summing over all n control points to obtain an expression for l_7, which we can then substitute back into (4). This provides a set of equations of the form:

$$l_1 \Delta x(i) + l_2 \Delta y(i) + l_3 \Delta x'(i) + l_4 \Delta y'(i) + l_5 \Delta x''(i) + l_6 \Delta y''(i) = 0 . \quad (5)$$

Where $\Delta x(i)$ is the distance of the control point from the centroid of the object in each image. There are now similar measurement errors on each term in (5) which may therefore be solved straightforwardly as a total least squares problem. To do so, we combine the equations for all n control points in matrix-vector form:

$$X\underline{l} = 0 \,, \tag{6}$$

where X is an n x 6 rectangular matrix, whose rows are equal to $(\Delta x(i), \Delta y(i), \ldots, \Delta y''(i))$.

The total least square solution which minimizes the sum of the squared errors on the co-ordinates $\Delta x(i)$, $\Delta y(i)$, $\Delta x'(i)$, $\Delta y'(i)$, $\Delta x''(i)$, $\Delta y''(i)$ is then obtained by a singular value decomposition (SVD) of X [18,19], and the singular vectors, $\underline{l}$ and $\underline{m}$, corresponding to the two smallest singular values of X provide, with the expressions for l_7 and m_7, the best two linear relationships between the image co-ordinates. We assume, for calculation of the distance measures described below in section 4, that the singular vectors are, as usual, normalized to unit length, $\|\underline{l}\|^2 = \|\underline{m}\|^2 = 1$

4 Evaluation

In the remainder of this paper, we assess the benefits of using the total least squares error approach described above, first by considering its geometric accuracy using a synthetic, test object, then by considering the reconstruction of computer generated images and finally by comparison of reconstructed images of views of real objects, using a laboratory test scene. A face image is also reconstructed to illustrate the process on more general, natural imagery. The last two comparisons require calculation of image intensities in addition to the co-ordinates (x,y), for which we follow an interpolation scheme similar to that proposed in [9-11].

In [9-11], image intensities were interpolated by using a measure of the distances, d' and d'', of the target image $I(x,y)$ from the basis images $I'(x',y')$ and $I''(x'',y'')$ respectively defined from the coefficients $a_1..a_4$ and $b_1..b_4$ in (3). We proceed similarly from our total least squares solution using the singular vectors $\underline{l}$ and $\underline{m}$, and define the *relative* image distances as:

$$d'^2 = l_5^2 + l_6^2 + m_5^2 + m_6^2 \text{ and } d''^2 = l_3^2 + l_4^2 + m_3^2 + m_4^2 \,. \tag{7}$$

As for the distance measures used in [9-11], d' correctly vanishes whenever the geometry of $I(x,y)$ may be obtained by a simple affine warping of $I'(x',y')$ alone, and similarly for d'' and I''. Moreover, unlike the scheme used in [9-11], we can, if desired use the symmetrical form of (5) to define a similar distance d between the basis views. These relative distance measures are sufficient for the image interpolation required here although, if desired, one could envisage making them absolute by, for example, averaging d over the ensemble of image views I of the object (or scene) of interest.

To interpolate I' and I'' to reconstruct I we then simply weight them accordingly

$$I = w'I' + w''I'' \,, \tag{8}$$

with weights

$$w' = d''^2 / \left(d'^2 + d''^2\right) \text{ and } w'' = d'^2 / \left(d'^2 + d''^2\right) \tag{9}$$

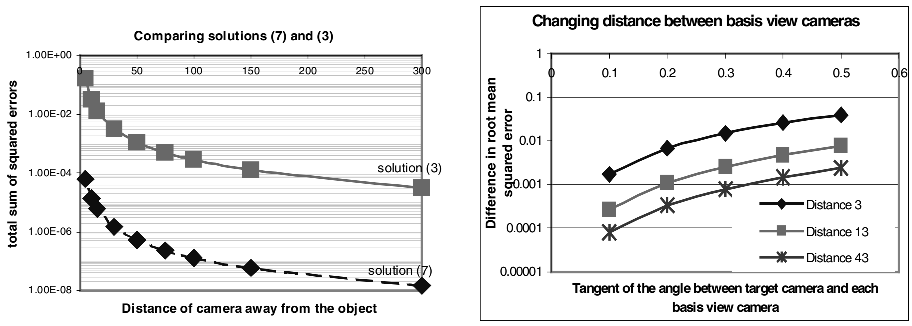
This defines the image intensity I at control points (x,y). In order to reconstruct I at all pixel locations, we triangulate the target image I and perform a piecewise linear mapping inside each triangle as described in [25] and used previously in [9-11]. Such an interpolation is consistent provided the image co-ordinates are linearly related and the triangulation is constrained to contain known object and surface boundaries. Although Delaunay triangulation algorithms have been developed [28] to take account of such constraints, in our examples, the triangulation, like the correspondence of the control points was performed manually.

4.1 Synthetically generated object

To compare the geometric accuracy of the total least squares solution (5) and the classical approach (3) used previously in [9-11], we used the vertices of a translucent geometrical test object in synthetically generated images as control points. The test object had ten vertices, eight at the corners of a unit cube and two at co-ordinates at (2,2,2) and (-2,-2,-2) on a diagonal line through opposite corners of the cube.

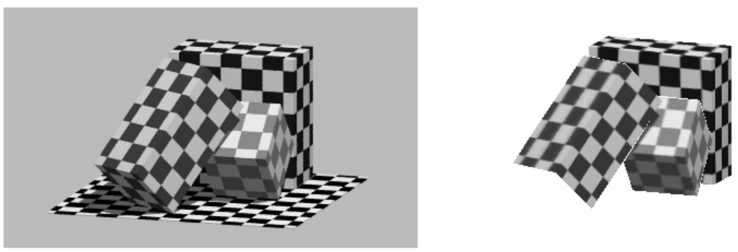
A series of tests were carried out on this object with the camera centres at several different positions. The tests included change of the distance of the cameras from the object, moving the target camera closer to one of the basis cameras, and moving the target camera perpendicular to the line joining the two basis camera centres. These tests were repeated 50 times with varying levels of random noise, between 0.26-2.5 pixels, added to *all* image control point co-ordinates.

To see how the solutions performed as the distance between the cameras and the object was varied, we first placed the target camera at (0,0,300) and the two basis cameras at (-30,0,300) and (30,0,300). We then moved all the cameras towards the object on a straight line from their initial positions 300 units from the object to within 5 units of its centre at (0,0,0). As shown in figure 1(a), the total sum of squared errors for the total least squares solution (5) is always considerably less than that obtained from the classical least squares solution (3). Figure 1(b) shows the difference in the root mean squared error of the two solutions, (3) – (5), as the basis view cameras were moved further apart. The target camera was kept halfway between the two basis cameras. The improvement factor obtained by use of the total least squares approach is, as might be expected, larger at smaller distances, but approximately independent of the angle between the basis views.



(a) (b)

Figure 1: Graphs comparing classical least squares and total least squares errors



(a) (b)

Figure 2: simulated boxes and reconstruction

4.2 Simulation

A set of images of a collection of coloured boxes similar to those used in section 4.3 below were simulated using Povray under orthographic projection. Under these circumstances and with no errors on the control points, a linear reconstruction is exact [1]. The reconstructed image is shown in figure 2(b). The first impression is of an accurate, but slightly blurred version of the target image 2(a). The blurring is a result of the intensity interpolation (8) and of the bilinear interpolation used to compute the basis image intensities I' and I'' between pixel co-ordinates. Similar effects have been reported by Pollard and he has suggested more sophisticated schemes that preferentially weight high spatial frequency components to the closer basis image [29]. There is also some aliasing at the boundaries of the boxes, but this could easily be corrected by weighting contributions to pixels on the boundaries appropriately.

4.3 Real Images

In order to use the relationships to reconstruct an actual image, a laboratory test scene was set up comprised of three boxes covered in coloured wrapping paper. A collection of photographs were taken at various, known camera positions, both of the boxes and of a calibration object made up from two planar tiles at an angle of 90^{o}, both similarly covered by colourful pictures.

(a) (b)

Figure 3: original and reconstruction of calibration targets

The corners of the boxes and tiles were used as control points, selected manually in all images, and used as input to the total least squares solution (5). These relationships were then used them to locate the positions of the control points in the third image using the known correspondence between the first two. The target image was manually triangulated so that all triangles fall on faces of the boxes or tiles. The control points were then projected into the third image using two equations of the form of (4). For the calibration tiles, figures 3(a) and (b) respectively show the actual target image, and the reconstruction. Apart from slight blurring, the reconstruction appears to be excellent.

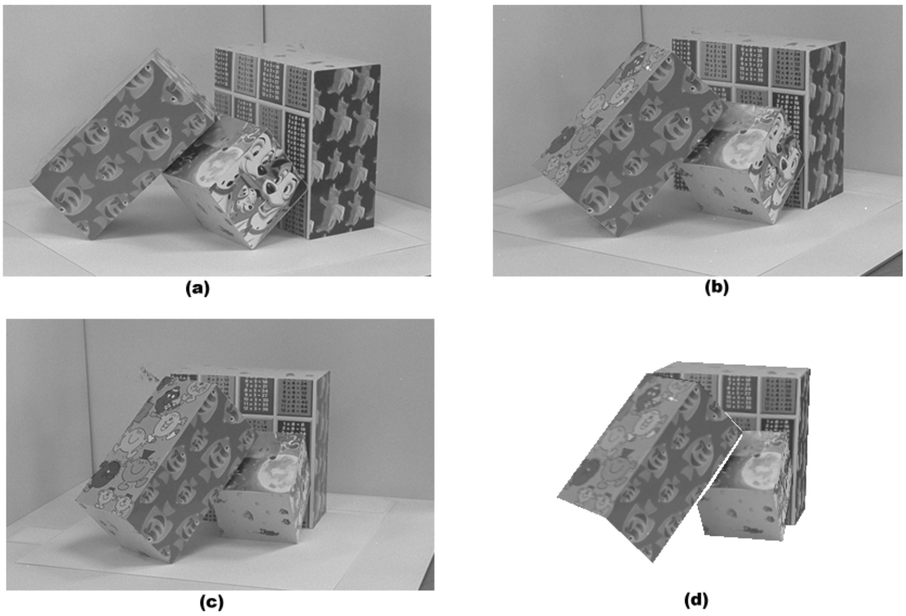


Figure 4: (a) and (b) basis views, (c) original target, (d) reconstruction

For the boxes, figures 4(a) and (b) were used as the basis images and 4(c) is the target image that we are trying to reconstruct. Again, the first impression (figure 4(d)) is of an accurate, but slightly blurred reconstruction, with some small, but noticeable artefacts at some of the occluding boundaries of the boxes. In particular, a small gap appears where the edge of the box nearest the cameras occludes the face of the smallest box. In this case it is not an aliasing effect. It is caused by the fact that the approximate linear mapping does not bring points in the target image exactly to their correct, corresponding positions. In practice, when the target image exists, as for example, in image encoding applications [10,11] the control point co-ordinates may themselves be used directly to define the triangulation in the target image and the subsequent interpolation, with the linear mapping entering only via the weights (9). In such cases, such problems caused by the approximate nature of the linear mappings are therefore circumvented, but in other applications, where the target image is not available, an alternative solution is required to ensure a consistent reconstruction. This is currently under investigation on a separate, but related research project aimed at using these techniques for the visualisation of historical artefacts over the Internet. Similar problems occur at other occluding boundaries where the pixels may overlap. Although not easily visible in figure 4(d), such overlap occurs at the top of the front box and has been highlighted by insertion of a red line. However, in this case it is easy to use the relative affine depth [2,30] to render only pixels on the nearest surface, essentially [9] by solving for the depth Z at each control point instead of eliminating it as in (3). Finally, we note that, with the present interpolation scheme, we have not reconstructed parts of the image present in only one basis view although, given the correspondences and triangulation, this should be easy to fix.

4.4 Face images

Finally, figure 5(b) and (c) show part of a face that has been reconstructed using the classical least squares (3) and the total least squares (5) schemes. The two relationships were estimated using the same set of control points. It can been seen that the two solutions are very similar and are both of very good quality. If inspected closely it can be seen that the shapes of the reconstructed sections are different, where the total least squares scheme has located the control points more accurately. There is also an improvement in the mapped intensities with the total least squares reconstruction but this is not visually obvious.

(a) (b) (c)

Figure 5: Reconstruction of a face image

5 Conclusions

It has been shown that a careful treatment of the measurement errors in a linear combination of views procedure by use of a total least squares solution leads to significant improvements in both the geometry and the appearance of the images generated. A synthetic object, simulated images produced by a ray tracer, laboratory test images of a simple scene, and a face image were used to assess performance of the approach. In particular, the synthetic object was used to show that the total least squares solution produces smaller geometric errors than a classical least squares solution. The simulated images, constructed under orthographic projection so that there were no geometric errors, were used to illustrate the quality of the image intensity interpolation. The laboratory test images were then used to show that the overall quality of reconstruction was similarly excellent for real data. They also indicated, however, that there were artefacts at occlusion boundaries caused by the approximate nature of the linear mapping between the images. Finally, face images were used to show that use of the total least squares procedure can lead to useful improvements in the quality and realism of the reconstruction of natural imagery.

In this paper we have only dealt with linear relationships which, under affine imaging conditions, are a good approximation to the full perspective trifocal tensor relationships. These linear relationships are, in fact, equivalent to the affine trifocal tensor described by Mendonca and Cipolla [31]. In the experiments reported where we were reconstructing an existing test image, a total least squares (TLS) solution [18] could be used since errors on the control points were independently and identically distributed. However, it was pointed out that in general when, say, a novel view of an object or scene is to be generated, the generalised total least squares (GTLS) procedure [22] should be used. Such a generalisation does not affect the principle of our approach, but would lead to a slight increase in the complexity of the solution since it requires a generalised singular value decomposition rather than the more familiar SVD [23]. Finally, we note that to apply either the total least squares or generalised total least squares procedure to the non-linear trifocal tensor equations would, in general, require an iterative solution.

Acknowledgement

This research was undertaken within the Postgraduate Training Partnership established between Sira Ltd and University College London. Postgraduate Training Partnerships are a joint initiative of the Department of Trade and Industry and the Engineering and Physical Sciences Research Council.

6 References

[1] Ullman, S. and Basri, R., “Recognition by Linear Combination of Models”, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 13, No 10, pp 992-1006, 1991.
[2] Shashua, A., “Photometry and Geometry in 3D Visual Recognition”, Ph.D Thesis, Massachusettes Institute of Technology, 1992.
[3] Chen, S.E., “View interpolation for image synthesis”, Proc. SIGGRAPH 93, in Computer Graphics, pp 279-288, 1993.
[4] Seitz, S.M and Dyer, C.R., “Physically-valid view synthesis by image interpolation”, Proc. IEEE Workshop on Representation of Visual Scenes, pp 18-25, 1995.
[5] Gortler, S.J., Grzeszczuk, R, Szeliski, R. and Cohen, M.F., “The Lumigraph”, Proc. SIGGRAPH 96, in Computer Graphics, pp 31-42, 1996.
[6] Avidan, S. and Shashua, A., “ Novel View Synthesis in Tensor space”, Proc. of Computer Vision and Pattern Recognition 1997, pp 1034-1040, 1997.
[7] Netzer, A. “Computer vision: Image based rendering”, notes from lectures presented by A Shashua, see http:// www.cs.technion.ac.il/~arnonn/IBR/introduction.html.
[8] Pollard, S., Pilu, M., Hayes, S., and Lorusso, A., “View Synthesis by Trinocular Edge Matching and Transfer”, in Proc BMVC’98, pp 770-779, 1998.
[9] Buxton, B.F., Shafi, Z. and Gilby, J., “Evaluation of the construction of novel views by a combination of basis views”, in Signal Processing IX, Theories and Applications, Proceedings of EUSIPCO98, the IX European Signal Processing Conference, Greece, Vol III, pp 1285-1288, Eurasip September 1998.
[10] Koufakis I. and Buxton, B.F., “Linear combination of face views for low bit rate face video compression”, in Signal Processing IX, Theories and Applications, Proceedings of EUSIPCO98, the IX European Signal Processing Conference, Greece, Vol IV, pp 2305-2308, Eurasip, September 1998.

[11] Koufakis, I. and Buxton, B.F., “Very low bit rate face video compression using linear combination of 2D face views and principal components analysis”, accepted for publication in Image and Vision Computing.
[12] Mundy J. and Zisserman A., “Geometric Invariance in Computer Vision”, MIT Press, Cambridge, Mass. 1992.
[13] Faugeras, O.D and Robert, L., “What can two images tell us about a third one?”, ECCV 94, pp 485-492, 1994.
[14] Torr, P.H.S. and Zisserman, A., “Robust Parameterization and Computation of the Trifocal Tensor”, BMVC 1996, Vol II, pp 655-664, 1996.
[15] Hartley, R., “Computation of the Quadrifocal Tensor”, In Proceedings of 5th European conference on Computer Vision, ECCV'98, Springer Verlag, pp 20-35.
[16] Heyden, A., “A Common Framework for Multiple View Tensors”, In Proceedings of 5th European conference on Computer Vision, ECCV'98, Springer Verlag, pp 3-19.
[17] Mühlich M. and Mester R. (1998) “The Role of Total Least Squares in Motion Analysis”, In Proceedings of 5th European conference on Computer Vision, ECCV'98, Springer Verlag, pp 305-321.
[18] Golub G. H. and Van Loan C. F. (1996) “Matrix Computations”, Third Edition, John Hopkins University Press.
[19] Lawson C. L. and Hanson R. J.(1995) “Solving Least Squares Problems”, Classics in Applied Mathematics, SIAM.
[20] Demmel J. W. (1987) “The smallest perturbation of a submatrix which lowers the rank and constrained total least squares problems”, SIAM Journal on Numerical Analysis, 24(1), pp 199-206.
[21] Pollard, S., Hayes, S., Pilu, M. and Lourusso, A., “Automatically synthesising virtual viewpoints by trinocular image interpolation”, HP Laboratories Technical Report, HPL-97-166, December, 1997.
[22] Van Huffel S. and Vandewalle J. (1989) “Analysis and properties of the generalized total least squares problem $AX \cong B$ when some or all columns in A are subject to error”, SIAM Journal of Matrix Analysis and Applications, 10(3), pp 294-315.
[23] Wilkinson, J. H., “The Algebraic Eigen Value Problem”, Clarendon, 1965.
[24] Trivedi, H., private communication, 1986.
[25] Goshtasby A. (1986) “Piecewise Linear Mapping Functions for Image Registration”, Pattern Recognition, 19(6), pp 459-466.
[26] Chen, S.E., “QuickTime VR - An Image-Based approach to Virtual Environments Navigation”, Proc. SIGGRAPH 95, in Computer Graphics, pp 29-38, 1995.
[27] Kanade, T., Narayanan. P.J. and Rander, P.W. “Virtualised Reality: Concepts and Early Results”, Proc. IEEE Workshop on Representation of Visual Scenes, pp 69-76, 1995.
[28] Pebay, P. P., “Consruction d’une constrainte Delaunay-admissible en dimension 2”, INRIA Research Report No 3492, September, 1998.
[29] Pollard, S. B., “View synthesis and creating immersive video object”, paper presented at the BMVA meeting on “Vision in the Media”, British Institute of Radiology, 10 March 1999.
[30] Ullman,S., “High Level Vision”, Cambridge, Mass. MIT Press, 1996.
[31] Mendonca, P.R.S. and Cipolla, R., “Analysis and computation of an affine trifocal tensor”, in Proc BMVC’98, pp 125-133, 1998.