

Simple Feature Identification and Matching for Stereoscopic Imaging

J Courtney[♠] and A M dePaor*

Dept. of Electronic and Electrical Engineering,
University College Dublin,
Belfield, Dublin 4, Rep. of Ireland.
E-mail: [♠]Jane.Courtney@ucd.ie
*Annraoi.dePaor@ucd.ie

Abstract – Many camera calibration techniques used in stereoscopic imaging require corresponding points in image pairs to be identified. The best matches tend to come from very distinctive regions, hence it is preferable to use only these points and reject points from more homogeneous regions. Often points are selected and matched by a combination of user input and automated matching. However, for most techniques, the more matching points identified, the more accurate and robust the calibration. Therefore, it is desirable to alleviate the need for user input and to automate the point selection process. In the feature matching technique described here, a combination of object extraction, feature identification and feature matching is used. This technique is simple, efficient and robust. The algorithm will be used as part of a larger project in 3D human motion capture which is currently under development at University College Dublin and the National Rehabilitation Hospital in Dun Laoghaire.

Keywords – Stereoscopic imaging, feature matching.

I INTRODUCTION

The first stage of stereoscopic imaging and subsequently 3D reconstruction is to describe the geometric relationship between a pair of stereo images. The image pairs are related by an *epipolar transformation* [2], which is determined by first computing the *fundamental matrix*. This is a 3x3 matrix which contains the geometric information relating stereo images. If \mathbf{p}_1 is a point in image 1 and \mathbf{p}_2 is its corresponding point in image 2, then the relationship between them is

$$\mathbf{p}_2^T \mathbf{F} \mathbf{p}_1 = 0 \quad (1)$$

where \mathbf{F} is the fundamental matrix. To compute the fundamental matrix, several matching points must be obtained, thereby forming several linearly independent equations for solving for the parameters of \mathbf{F} [3, 4]. According to [5], \mathbf{F} is homogeneous, of rank two and has seven independent parameters and

so requires at least seven matched points to be computed accurately. However, with just seven points, there are three possible solutions for the parameters of \mathbf{F} [4]. Also, each match must be guaranteed accurate to avoid errors in the calculations. By using many more than seven points to determine the parameters of \mathbf{F} , we not only ensure the uniqueness of the solution, but also make our calculations less sensitive to erroneous matches [1].

In order to begin computing the fundamental matrix, an automated matching technique must first be devised which is efficient enough to match many points but robust enough to guarantee a high percentage of correct matches. The best matches occur at clearly identifiable regions or ‘features’. Thus, in an automated point selection system, these points are preferable to points lying in homogeneous regions. The technique described here isolates the area of interest in image 1 (usually an object, human or otherwise), then identifies feature points within the area of interest and finally matches the selected feature points to find their counterparts in image 2.

II OBJECT IDENTIFICATION

The first step in selecting points of interest is to identify the area of interest within the image. The images used here have been recorded in a controlled environment with an uncluttered, dark background and invariant lighting. However, similar images can be achieved by various foreground identification methods, e.g. background subtraction, e.g. [6], or foreground likelihood imaging [7]. Once the background has been sufficiently segmented from the object in the image, the area of most interest can be defined by identifying the mean-crossing points. These are defined in the x-direction as the points between which the vertical projection of the intensity, $V(x) = \sum_{y=1}^h I(x, y)$ is greater than the mean over x as defined by

$$\bar{V} = \frac{1}{w} \sum_{x=1}^w V(x) \quad (2)$$

where w is the width of the image and h is its height. Similarly, in the y-direction the horizontal projection is given by $H(x) = \sum_{x=1}^w I(x, y)$ and its mean is

$$\bar{H} = \frac{1}{h} \sum_{y=1}^h H(y) \quad (3)$$

The points where these means are crossed define the boundaries of the region of interest. Once the region of interest is defined, further calculations can be limited to this region, thereby making the algorithm more efficient.

III FEATURE RECOGNITION

Once the area of interest has been determined, points must be chosen within this region and tested for their homogeneity. A grid of points is selected with points at fixed intervals. The intervals should be sufficiently large to minimise calculations but small enough to allow matching of plenty of points.

At each of these points the suitability is tested using a simple Sobel edge mask:

$$\mathbf{M} = \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix}$$

The point being tested and its eight neighbours are labelled as follows

$$\mathbf{X} = \begin{pmatrix} X_0 & X_1 & X_2 \\ X_3 & X_4 & X_5 \\ X_6 & X_7 & X_8 \end{pmatrix}$$

(X_4 is the point being tested) and the parameters of \mathbf{M} are similarly indexed. An edge value, E , can then be assigned to the point:

$$E = \sqrt{\left(\sum_i M_i \cdot X_i\right)^2 + \left(\sum_i M_i^T X_i\right)^2} \quad (4)$$

In calculating the edge value, the colour information of the pixels is preserved. This means that the variation across all colours is tested so that no feature is missed. More sophisticated approaches may be used to devise a similar test function (e.g. by testing the variance over the region or corner matching), however, the Sobel masks' simplicity and reliability adds to the robustness and speed of this method and a high value for E is sufficient proof that the point is an element of a feature.

IV FEATURE MATCHING

The final stage of the algorithm involves matching the features selected in image 1 to their counterparts in image 2. This is achieved using a least squares matching technique with the normalised correlation coefficient as a measure of match.

A template patch is selected about the point of interest in image 1 and its best match in image 2 is found by searching for this template in a defined window. At this stage, a large variety within the template is important to ensure a unique match. Also, dimensions of the template and search window are very important [8].

A large template will give more detail but will be difficult to match exactly without including rotations and affine warping [9] and thereby slowing down the process. A small template will be less detailed and so will give several erroneous matches in the second image. Similarly, the size of the search window will effect the accuracy of the match. A small window is preferable to avoid erroneous matches but the minimum allowable size of the window will depend on the variation between image 1 and image 2. The effect of the size of the windows on the quality and accuracy of the results and on the speed of the matching algorithm has been shown in [8] and complete results are given.

At each point in the search window a test patch of pixels, \mathbf{S} , is obtained and this patch is compared to the pixels of the template patch, \mathbf{T} (See Figure 1).

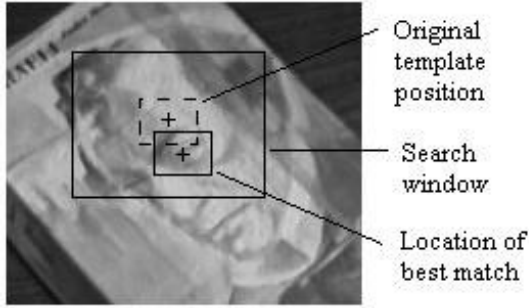


Figure 1: Template matching

The normalised correlation coefficient (5) is used as a measure of match.

$$\text{corr}(\mathbf{S}, \mathbf{T}) = \frac{N \sum_{i=1}^N S_i T_i - \sum_{i=1}^N S_i \sum_{i=1}^N T_i}{\sigma_s \sigma_t} \quad (5)$$

where N is the number of pixels in the patch and σ_s and σ_t are the standard deviations of patch \mathbf{S} and template \mathbf{T} respectively. When calculating the normalised correlation coefficient, the colour information of the pixels is preserved to increase the accuracy of the match.

Rotating the patch as well as translating it through the search window can achieve a more accurate match but this greatly increases the search space and therefore the computational costs of the algorithm. A more elegant algorithm for locating the best position *and* orientation has been proposed by [10]. This is achieved by examining the intensity changes within the initial approximation patch (found as described) and using these to solve for rotational and translational parameters. The more skewed the camera's viewpoints are from each other, the more important this rotational element becomes. However, in stereoscopic imaging, the cameras are usually side-by-side or very close and so simple template matching is often sufficient.

V FUNDAMENTAL MATRIX CALCULATION

Using the feature points that have been identified and matched using this simple algorithm, parameters for the fundamental matrix can then be calculated. From equation 1, we have a relationship between the identified points and their matched counterparts. This equation can be expanded to give the matrix equation

$$\begin{pmatrix} x_2 & y_2 & 1 \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} = 0 \quad (6)$$

where (x_1, y_1) and (x_2, y_2) are the co-ordinates of a point in image 1 and its match in image 2 respectively and the f_{ij} are the elements of the fundamental matrix, \mathbf{F} . This matrix equation can be re-written in the form

$$\begin{bmatrix} \mathbf{p}_2^T \otimes \mathbf{p}_1^T \end{bmatrix} \mathbf{f} = 0 \quad (7)$$

where $\mathbf{p}_2^T \otimes \mathbf{p}_1^T$ is the *Kronecker Product* of \mathbf{p}_2^T and \mathbf{p}_1^T , i.e. each element of \mathbf{p}_2^T is replaced by that element times \mathbf{p}_1^T . The 9×1 column vector \mathbf{f} is the column of the columns of the 3×3 matrix \mathbf{F}^T i.e., written as a row vector it becomes:
 $\mathbf{f}^T = (f_{11} \ f_{12} \ f_{13} \ f_{21} \ f_{22} \ f_{23} \ f_{31} \ f_{32} \ f_{33})$.

Finally, by expanding equation (7) to incorporate the n matched points provided by our algorithm, we get a set of n equations that can be solved for the parameters of \mathbf{F} :

$$\begin{pmatrix} x_{11}x_{21} & x_{21}y_{11} & x_{21} & x_{11}y_{21} & y_{11}y_{21} & y_{21} & x_{11} & y_{11} & 1 \\ x_{12}x_{22} & x_{22}y_{12} & x_{22} & x_{12}y_{22} & y_{12}y_{22} & y_{22} & x_{12} & y_{12} & 1 \\ & & & & & & & & \vdots \\ x_{1n}x_{2n} & x_{2n}y_{1n} & x_{2n} & x_{1n}y_{2n} & y_{1n}y_{2n} & y_{2n} & x_{1n} & y_{1n} & 1 \end{pmatrix} \mathbf{f} = 0 \quad \dots(8)$$

VI RESULTS

A pair of stereoscopic images of a simple object with a homogeneous background was used as an initial test input for the program (see figure 2). These test images were taken from the stereoscopic image library on the website of the Vision and Autonomous Systems Center of the Robotics Institute of Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Although there are some erroneous matches, these are outweighed by many strong and accurate matches. The matched point sets from these images were used to calculate the fundamental matrix which was then used to determine a *perspective projection matrix*. This is a matrix which maps the 2D image points to the 3D world axis. This mapping allows the 2D image of the object to be reconstructed in the 3D world. The result of the 3D reconstruction for the test images is shown in figure 3.

The method was also applied to a selection of more complex objects including human limbs, which this system will ultimately be used to model. Both the simple test images and the human limb images give

good results, however, 3D reconstruction of human limbs is a more complex procedure and will be dealt with in future work.

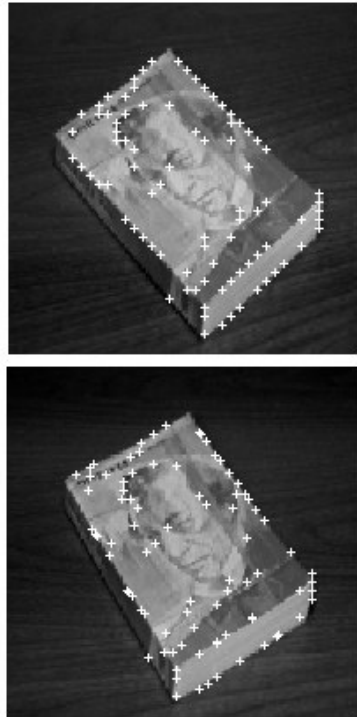
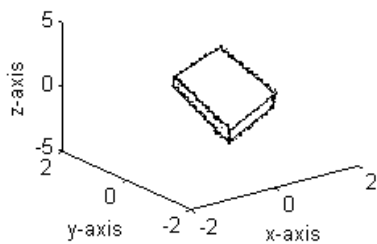
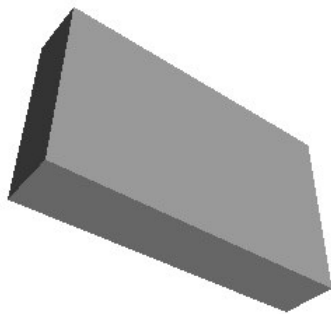


Figure 2: Matched image points on the test images.



(a)



(b)

Figure 3: Reconstruction of the object in the 3D world: (a) MATLAB 3D plot; (b) OpenGL object.

REFERENCES

- [1] B. Boufama, R. Mohr. "Epipole and fundamental matrix estimation using virtual parallax." *Proceedings of the Fifth IEEE International Conference on Computer Vision*. 1030-1036, June 1995.
- [2] O.D. Faugeras, Q.T. Luong, S.J. Maybank. "Camera self-calibration: theory and experiments." *Proceedings of the Second European Conference on Computer Vision*. 321-334, May 1992.
- [3] O. Hesse. "Die cubische Gleichung von welcher die Lösung des Problems der Homographie von M. Chasles abhängt." *J. reine angew. Math.*, 62:188-192, 1863.
- [4] R. Sturm. "Das Problem der Projektivität und seine Anwendung auf die Flächen zweiten Grades." *Math. Ann.*, 1:533-574, 1869.
- [5] S. Maybank. *Theory of Reconstruction from Image Motion*. Springer-Verlag, 1993.
- [6] E. Hayman, J.O. Eklundh. "Statistical background subtraction for a mobile observer." *Proceedings of the Ninth IEEE International Conference on Computer Vision*. 67-74, Oct 2003.
- [7] Halvorsen, K, "Model-based methods for tracking and analysis of human movement". *Licentiate Thesis*, Uppsala University, 1999.
- [8] M. Betke, J. Gips, P. Fleming. "The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 10:1-10, March 2002.
- [9] N. D'apuzzo. "Automated photogrammetric measurement of human faces." *International Archives of Photogrammetry and Remote Sensing*. 32:402-407, 1998.
- [10] S.Y. Zheltov, A.V. Sibiryakov. "Adaptive subpixel cross-correlation in a point correspondence problem." *Optical 3D Measurement Techniques*, 86-95, September 1997.