Asian Journal of Information Technology 5 (3) : 253-257, 2006

3D Shape Extraction from Uncalibrated Environments and Video Camera

[1,2]A.S.T. Hussain, [1]N.E. Berrached, [2]A.E. Murad and [2]Tayeb Basta
[1]University of Science and Technology of Oran (USTO), Oran 31000, Algeria
[2]University of Sharjah, Sharjah, UAE

Abstract: In this study, we have described an approach for a 3D scene reconstruction using 2 randomly selected adjacent colored video frames. We have used a single uncalibrated video camera to take a record for uncalibrated environment. The Selection of 2 frames based on the maximum homogeneity of points on these frames are favorable; could be any two adjacent frames. The use of Harris technique were very useful to find the edges and corners on each selected image (frame), then the use of the autocorrelation function based on Gaussian's function been used to find the corresponding matched points. Then the correlated matched pair points are found on both images and by calculating the gradient of the correlated paired points on both images represents approximately the Z direction (calculating dzdx and dzdy). This is yielding that each point on each image (frame) can be represented in a 3D coordinates which yields to 3D shape estimation, which is achieved by the RANSAC function. This method has got an errors, because of its dependence on probability. The main advantages of the proposed approach is applicable for indoor and outdoor application. This technique is suiting the natural real world applications. The proposed method is illustrated on a set of examples of an indoor captured colored video records, the selected frames are selected randomly from the video records.

Key words: 3D, reconstruction, autocorrelation gaussian RANSAC

INTRODUCTION

Reconstructing the scene from image sequences captured by moving cameras with varying intrinsic parameters is one of the major achievements of computer vision research in recent years.

The reconstruction process is a set of consecutive tasks to be accomplished. First task consists of extraction of point landmarks, lines/curves and surfaces/regions in 2D. Secondly, regarding the geometry used; two frames means two-views, relevant methods for determining the correspondence between features are used. Then, appropriate transformation functions for image registration are applied. Finally, determination of the reliability, accuracy and speed of applied image registration methods are verified and compared to their counterpart available in the literature.

Here, we are presenting a light survey in which the problem of 3D extraction from a frame sequences is tackled. After that, we investigate the literature for the use of two-views to extract 3D structure. The basic geometrical and algebraic notions that make the foundation for such extractions follow the latter investigation. After that, we shade some light on different image registration techniques and finally we quote some words about camera self-calibration (Uncalibrated) techniques as we think that it can be exploited somehow to reach our objective in many different ways[1-5].

PREVIOUS WORKS OF 3D EXTRACTION

In the challenge of generating a qualitatively accurate 3D reconstruction of the actions performed by an individual in a long (30 seconds) monocular image sequence[1,6,7,8]. It is assumed the individual is not wearing any special reflective markers or clothing. Any solution must be able to cope with the multitude of difficulties that may arise over several concurrent frames: severe self-occlusion, unreliability of methods for limb and joint detection, motion blur and the inherent ambiguities in reconstructing rigid links from monocular images. Until now, the only approach guaranteed to produce a complete and accurate reconstruction in such circumstances is: for each frame in the sequence, manually locate the skeletal joints and perform 3D reconstruction using the method of Mohan *et al.*[2]. The latter involves solving the forward/backward binary ambiguity for each rigid link by inspection and estimating the relative lengths of each limb. For very short sequences, this is a relatively painless procedure, but rapidly becomes impractical for longer sequences.

The traditional tracking approach to human motion capture[9] is to perform manual initialization at the beginning of the sequence and then update the estimate of the reconstruction over time in accordance with the incoming data.

Corresponding Author: A.S.T. Hussain, University of Science and Technology of Oran (USTO), Oran 31000, Algeria

In[10], the authors considered the entire sequence, approximate the actions present by a set of representative frames (automatically determined from the sequence) and from these obtained a coarse description of the subject's motion. Finer detail is added by locating the skeletal joints in each frame by extrapolating from manually initialized joint locations on the representative frames.

Their algorithm consisted of initialing extraction of the keyframes summarizing the sequence, the labeling of skeletal joint positions, formation of 3D keyframes and interpolation of 3D keyframes and then the final 3D reconstruction.

The authors presented the performance of their system for over 36 sec of tennis footage. They claimed that this is the longest full-body 3D reconstruction attempted from monocular image data[11-14].

Once the mobile robot understands the required job to be done, there is another problem which is the exact way of handling the object; the three dimensions of the object is needed to be extracted. In our application we use a 2D optical images, this lead us to extract the 3D of the object. Once the 3D shape extracted, the mobile can handles the object in a better way.

For more reliable applications, the robot is fitted with a video wireless camera. For example, the patient can follow the robot navigation in more details as well as handling the required object in more accurate way.

3D SHAPE EXTRACTION FROM A STEREO VIEW

There are two ways of extracting three-dimensional structure from a pair of images or a pair of video frames sequence. In the first and classical methods, known as the calibrated route, in which firstly need to calibrate either cameras (or viewpoints) with respect to some world coordinate system, calculate the so-called epipolar geometry by extracting the essential matrix of the system and from this compute the three-dimensional Euclidean structure of the imaged scene. The corresponding pixel view point of X is given by Eq. 1.

$$X' = R_x+\tau \quad (1)$$

Where X' is the corresponding view of X pixel point on the second image on the stereo pair, (R) is the rotation the two camera coordinate systems with two views and a translation (τ), (Fig. 1).

Taking the vector product with (τ), followed by the scalar product with (X') we obtain

$$X'.(\tau \wedge RX) = 0 \quad (2)$$

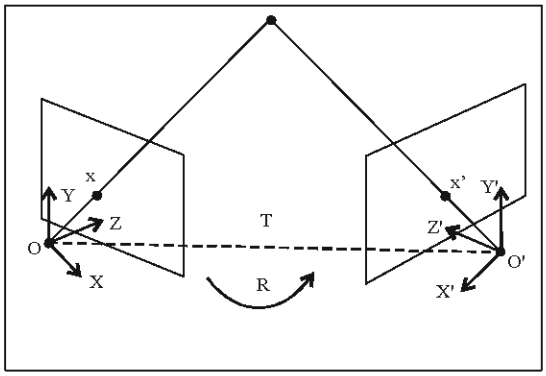

Fig. 1: The Euclidean relationship between the two view-centered coordinate systems

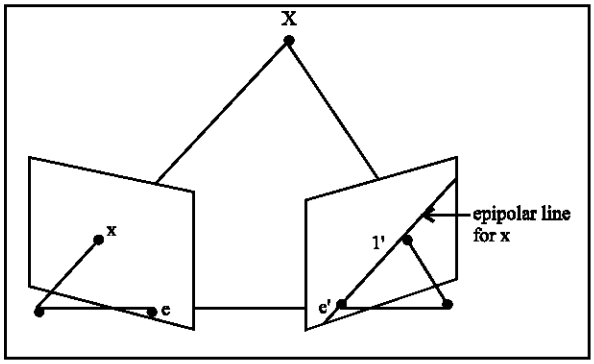

Fig. 2: The epipolar line along which the corresponding point for must lie (single point demonstration)

Which expresses the fact that the vectors (Cx), (C'x') and (C'C) are coplanar. This can also be written as

$$x'^T E = 0 \quad (3)$$

Where E is the essential matrix which can be written as follows in Eq. 4

$$E = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}.R \quad (4)$$

Where (τ) is represented as follow $\tau = (t_x, t_y, t_z)^T$. Equation (4) is the algebraic representation of epipolar geometry for known calibration and the essential matrix relates corresponding image points expressed in the camera coordinate system.

However, the second or uncalibrated route, the one we follow is more likely corresponds to real world applications, which we are using it in our application. In

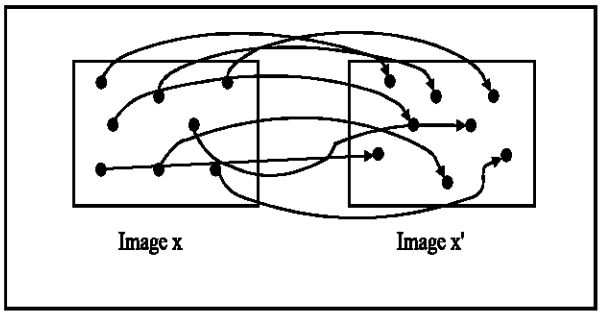

Fig. 3: The epipolar line along which the corresponding point for must lie (a set of matched points demonstration)

an uncalibrated system, a quantity known as the fundamental matrix, which is calculated from image correspondences (Fig. 2) and this is then used to determine the projective three-dimensional structure object on the imaged scene[9].

Making extensive use of Projective Geometry, several techniques have appeared in the last decades, which are capable of recovering the complete three-dimensional information from perspective images. An important feature of these algorithms is that no knowledge of the camera's positions and internal parameters is required: They can use uncalibrated images and produce so called projective reconstructions. For the case of stereo given images, all projectively relevant camera parameters are encapsulated in a single mathematical object, the stereo tensor. The tensor is all what is needed for projective reconstruction and can be estimated from image measurements (correspondences) alone as Fig. 3. However, reliable estimation of the stereo tensor is crucial for 3D reconstruction from uncalibrated cameras. Linear estimation is possible but not satisfactory because it does not enforce the nonlinear constraints that must be fulfilled by a valid tensor[10].

It is a condition for the fundamental matrix (f) to be of rank 2 and satisfy Eq. 5,

$$[x \quad y \quad 1]f\begin{bmatrix} x` \\ y` \\ 1 \end{bmatrix} = 0 \qquad (5)$$

Where (x,y) point on left image and (x`, y`) point on right image, also can be written as follows in Eq. 6 and more simplified in Eq. 7,

$$[x \quad Y \quad 1]^*\begin{bmatrix} f & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} * \begin{bmatrix} x` \\ y` \\ 1 \end{bmatrix} = 0 \qquad (6)$$

$$[x`x \;\; x`y \;\; x \;\; y`x \;\; y`y \;\; y \;\; x` \;\; y` \;\; 1]^*f = 0 \qquad (7)$$

Where f =[f1 f2 f3 f4 f5 f6 f7 f8 f9].

REAL WORLD APPLICATION

In our application; we have used a single uncalibrated video camera, take a record for uncalibrated environment for a period of time. It is preferable the used video camera has a high rate of frames recording per second, that yield to a sharp image will be represent in each frame. This technique is applicable for indoor and outdoor application. This technique is suiting the natural real world application.

The technique in steps:

- Take a record of few seconds or few minutes by using digital video camera.
- Select 2 frames based on the maximum homogeneity of points on these 2 frames; could be first and second, third, forth, or tenth.
- By using Harris technique to find the edges and corners on each selected image (frame), then use autocorrelation function based on Gaussian's function to find the corresponding matched points, (this method has got a huge error), because of its depends of probability.
- Once the correlated matched points are found on both images, then the gradient of the correlated paired points on both images represents approximately the Z direction (dzdx and dzdy), which is achieved by RANSAC function.
- Then in this case the correlated matched paired points can be represented in XYZ coordinates, then each point on each image can be represented in 3D coordinates which yields to 3D shape is estimation.

A RANSAC based procedure is described for detecting inliers corresponding to multiple models in a given set of data points. The algorithm we present in this study (called multiRANSAC) on average performs better than traditional approaches based on the sequential application of a standard RANSAC algorithm followed by the removal of the detected set of inliers. We illustrate the effectiveness of our approach on a synthetic example and apply it to the problem of identifying multiple world planes in pairs of images containing dominant planar structures[17].

STEPS IN DETAILS

I took a video film of 20 seconds Fig. 4, which is my own table, books and computer. This film taken at a rate of 30 frames per second.

Two frames are selected; fame number 146 and 151, Fig. 4, the selected frames depends on a difference between them not less than 3%.

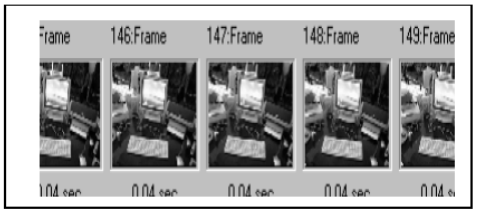

Fig. 4: A sample of some frames selected from a digital uncalibrated video camera

(A)

(B)

Fig. 5: The corresponding matched pair points on both selected frames (A and B)

Fig. 6: Shows both selected frames are coincides on top each other

By using Harris approach, finding the edge and corners on both frames. The use of autocorrelation and Gaussian's function is used to find the corresponding matched pair points on both selected frames (Fig. 5).

Fig. 7: Errors generated by Gaussians function when it is applied to all points in both frames

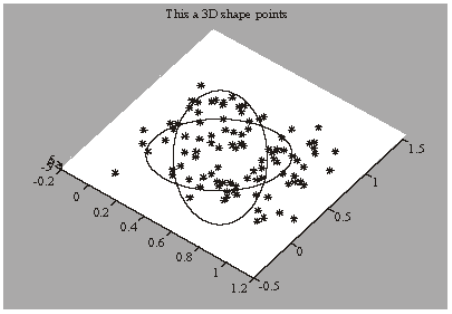

Fig. 8: 3D shape shows 3D correlated matched points

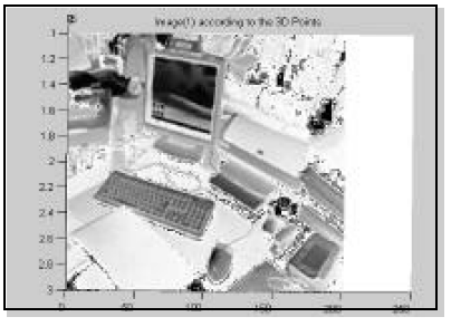

Fig. 9: Shows the estimated 3D image of the original 2D image

The homogeneity correlated points Fig. 5, obtained by using autocorrelation principles and a Gaussians function. Figure 6 shows both selected frames are coincides on top each other with a shown displacement difference between the correlated pair of points. This difference is used to calculate dzdx and dzdy which is used to estimate the Z direction.

Joining the homogeneity correlated points by straight lines, Fig. 7, long white lines are the errors generated by Gaussians function when it is applied to all points in both frames.

The fundamental and projective matrices are obtained based on geometrical calculated on both frames.

Finally a 3D shape represented by the following correlated matched points, shown on this Fig. 8, which are 3D coordinate points (XYZ) coordinates. The elliptical curves are added to Figure 8 shows these point are in a 3D representation.

Finally when we replace the texture of the original shape (which is represented by one of the frames) onto these 3D points, we get this shape (Fig. 9).

CONCLUSION

The main great problem was the error generated by using the auto correlation function which is based on Gaussians function to get the matching points on both selected frames. In this application, we have obtained 3D shape from a video film. We have used Harris principle to detect corners and edges in both frames from the taken video film. The second step was found out the matching points between the detected pair of points on both selected frames, on this step of matching means that the camera calibration is done. This calibration is done with an errors (when we say an errors, means that; a texture which is the image of one of the selected frames is stretched over the found matched pair points, but now these points have a coordinates of XYZ) that's yields to generate the fundamental and projected matrix, which is done.

The Z direction is obtained from the gradients of each pair correspondent points on both frames. The values of the gradient dzdx and dzdy represents the values on Z direction of each pixel, which is achieved by the RANSAC function.

The main great problem was the error generated by using the auto correlation function which is based of Gaussians function to get the matching points on both selected frames.

Nevertheless, we presented the mathematical foundation for the methods used in many different ways to reach good approximations for matrices that relates of 2D to 3D images.

The problem under study is amongst the few problems in the computer science field that are solved by well mathematically founded methods. Even though, we remarked the absence from the literature of reporting execution times or storage space required by algorithms used in the extraction of 3D structures.

ACKNOWLEDGEMENTS

I would like to thank Prof. N.E. Berrached (USTO) for his supervising throughout this research project, my thanks to Dr. Tayeb Basta and Mr. Ali Emad (University of Sharjah) for their assistance in terms of verifications and applications.

REFERENCES

1. Sujit Kuthirummal, C.V. Jawahar and P.J. Narayanan. Frame alignment using multiview constraints. http://www1.cs.columbia.edu/~sujit/Papers/Html/FAlignNcc02.htm
2. Mohan, Obeysekera. Affine Reconstruction from multiple views using Singular Value Decomposition. http://www.csse.uwa.edu.au/~pk/studentprojects/mohan/documentation/thesis/handed_cshonours.pdf
3. Epipolar geometry. From Computer Vision IT412 http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT10/node3.html
4. Byong, Mok Oh, Max Chen, Julie Dorsey and Frédo Durand, 2001. Image-Based Modeling and Photo Editing. http://graphics.lcs.mit.edu/ibedit/ibedit_s2001_cameraReady.pdf
5. Hussain, A.S.T. and N.E. Berrached, 2005. A3D Pattern Recognition based on fourier descriptors and neural networks@, ICMSAO/05, First International Conference on Modeling, Simulation and Applied Optimization, American University Of Sharjah, Sharjah, UAE.
6. Faugeras, O. and Q. Luong, 2001. The Geometry of Multiple Images. MIT Press.
7. Hartley, R. and A. Zisserman 2000. Multiple View Gemoetry in Computer Vision. Cambridge University Press.
8. Longuet-Higgins, H.C., 1981. A Computer Algorithm for Reconstructing a Scene from two Projections. Nature, 293: 133-135.
9. Szeliski, R. and P.H.S. Torr, 1998. Geometrically constrained structure from motion: Points on planes. In 3D Structure from Multiple Images of Large-Scale Environments, European Workshop SMILE'98, Lecture Notes in Computer Sci., 1506, pp: 171-186.
10. Gareth Loy, Martin Eriksson, Josephine Sullivan and Stefan Carlsson,. Monocular 3D Reconstruction of Human Motion in Long Action Sequences. http://eprints.pascal-network.org/archive/00000496/01/loy_eccv04.pdf
11. Martin Neil Armstrong, 1996. Self-Calibration from Image Sequences. PhD Thesis, Department of Engineering Science, University of Oxford, Michaelmas Term.
12. Maybank, S. and O. Faugeras, 1992. A theory of self-calibration of a moving camera. Intl. J. Computer Vision, 8: 123B151.
13. Zisserman, A., D. Liebowitz and M. Armstrong, 1998. Resolving ambiguities in auto-calibration. Phil. Trans. Royal Soc. London A, 356: 1193-1211.
14. Armstrong, M., A. Zisserman and R. Hartley, 1996. Selfcalibration from image triplets. In Proc. 4th European Conf. on Computer Vision, Lecture Notes in Computer Sci., pp: 3-16.