OBJECT RECONSTRUCTION FROM IMAGES OF A MOVING CAMERA

Jussi Heikkinen

Institute of Photogrammetry and Remote Sensing
Helsinki University of Technology
Finland
e-mail: Jussi.Heikkinen@hut.fi

Commission V, IWG V/III

KEY WORDS: feature based photogrammetry, feature matching, edge detection, hough transformation

ABSTRACT

This paper depicts a proposed method to model three dimensional objects from video sequence. The method is based on robust 2D features extraction, feature matching, and feature based photogrammetric modeling. The procedure to be presented can be considered as semiautomatic as its nature. During the whole process the operator will assist the system, but allowing the system itself to do all heavy routine tasks. The main idea is that we abandon using a point to point correspondence in object reconstruction and instead use a point to feature correspondence. In this case, a feature denotes a three dimensional object presented with a variable number of parameters. The parameters indicate the shape, size and position of object in selected coordinate system. This means that no form fitting is required in image space. The image observations resulted from edge detection can be used directly to estimate three dimensional object parameters.

1. INTRODUCTION

The idea of this procedure is to use all possible data of multiple video frames to solve the feature triangulation problem. The approach requires massive computation, that's why much of the low level feature extraction is executed as a background process. Using video imaging means an enormous amount of frames even though the recording session lasts only few minutes. This leads to a storing problem of the video sequence images. By using compression and processing data in smaller sequences this problem can be overcome.

CCD cameras have developed in few years and their stability has improved dramatically. Still, they do not reach the level of close-range analog photogrammetric cameras both in geometry and recording the brightness and contrast differences. The lack of precision of the camera can although be compensated by adding the amount of data. When using LSQ-type estimation the external accuracy is depending on stochastic properties as well as geometry. By adding observations the effect of noise can be diminished. The easiness and inexpensiveness of digital recording gives us this possibility to add the number of images and through this the number of observations. Naturally, cameras are supposed to be calibrated before the recording session.

Automatization of feature extraction needs robust algorithms to succeed. Insensitivity of algorithms against noise and poor contrast is required. The edge detection is dealt in Chapter 2. The line features are used as main features as they are easier to identify. For pixel grouping Hough transformation is implied. Although line extraction is executed in a batch job and is not as time restricted as in on-line systems, we are trying to speed up the process due to the amount of data. We are using the RHT-algorithm (Random Hough Transformation) for finding line features on image. This algorithm is based on random sampling and convergence mapping cycles succeeding each other. The method does not need as huge storage resources and work load as other stochastic Hough transformations but is still able to localize lines accurately. The RHT-algorithm is depicted more precisely in Chapter 3.

In Chapter 4 problems of finding the corresponding lines in succeeding frames will be considered. The low level feature extraction is done in the most automated way. In this paper we are not trying to present a fully automated measuring system, but to apply automatic methods in areas where tasks are easy to automate and need lot of routine work. Operator is used here as a guidance help. Operator's duty is to pick up all interesting lines on one image and add some constraints like parallelism, intersection of 3D-lines, perpendicularity etc. The automatic feature matching procedure will do the rest.

The 3D-feature reconstruction is based on methods presented by Mulawa and Mikhail[1,2], namely linear feature based photogrammetry. The approach aims to use liner features as primitive, so the point to point correspondence is substituted with feature to feature correspondence. That means we do not have to measure image points of a 3D-point, but image points belonging to a 3D-feature. The feature, in our case line feature, is identified in multiple frames and bundle adjustment is performed to solve the camera orientations as well as the feature line parameters. The feature lines are presented as parametric lines. This formulation has been used earlier in presenting geometrical modeling systems in CAD/CAM applications. The photogrammetric presentation binds one single pixel observation to 3D-feature estimation, so no line form fitting is needed in 2D space. The line features are not the only feature type which have this kind of formulation to present. Also circles, ellipses, and other conic section curves, as well as splines, have the photogrammetric relation between image observations and feature parameters. The lines are though the most robust features to identify and to extract. Actually, lines are not able to solve the feature triangulation alone, also other feature types like circles have to be included, unless some constraints of line intersection are determined. In this case, triangulation is possible with line features alone.

The feature triangulation constructs a stable frame work for additional measurements. Other features can be measured from image sequence by directing the measurements and identifying the feature types by the operator. After that, the automatic feature extraction as well as feature matching will do the rest; find and extract the feature from subsequent images.

This presented system can be adapted when measuring facades of buildings as well as other objects which have to be measured precisely. Also e.g. in car collision tests, the system can be applied in a little modified form.

2. IMAGE OBSERVATIONS

The three dimensional form fitting can be done by using pixel coordinates as observations, as well as points measured with subpixel accuracy. Using rough pixel coordinates means that we have larger variance of the observations but if the estimates are unbiased, results should be the same as when using subpixel coordinate values. The reliability is based on how accurately the edge detector can find the real edge and how invariant it is against noise of the image. The result of the LSQ-estimation is depending on the "goodness" of the observations. That means all gross errors have to be excluded out of the estimation with some robust way. One way of doing it is to use Hough transformation which is a very robust method to find out gross errors and to use it for feature classification. In this research we have chosen the *Random Hough Transformation* because of its low computing consumption and high accuracy. More details of RHT are given in Chapter 3.1.

2.1 Edge detection

For edge detection traditional gradient operators (Roberts, Prewitt, Sobel etc.) are adequate if the images are free of noise. In case of video images and outdoor circumstances, noise unfortunately is part of the game. Those gradient operators which indicate the local gradients, produce a large response for a large grayscale gradient, where the "Gradient-sum", proposed by Rosenfeld[3], is more immune to large edge spikes due to the smoothing effect of the summation. The standard procedure in practice is that before edge detection some smoothing will be performed for the image. The Canny operator[4] is based on linear gradient of the input signal with Gaussian smoothing as an integral part of the operator. This operator is appropriate for video images which usually need smoothing before edge detection. With Canny operator the level of smoothing is determined by the σ of the Gaussian function. The Canny operator is related to the Laplacian of Gaussian (LoG) operator, but it uses the first derivate of the Gaussian function when LoG uses the second derivate. The direction of the gradient can also be calculated, which might be helpful in the feature matching stage although the estimates of the direction are not quite accurate.

Edge strengthening is especially worthy when using noisy images. This helps finding the final edges by detecting the maximum of gradients. The automatic thresholding is often based on the maximum value and variance of gradients. The thresholding is applied for extracting all prominent edges and ignoring weak, noisy edges. Thresholding can be done locally or globally. Local thresholding means that in a smaller region the maximum gradient and the variance are calculated and in this area the threshold value is based on those indicators.

3. EDGE GROUPING

After finding the prominent edges, edges have to be grouped together with some criteria. That might be e.g. the common gradient direction. One way is to use a line following algorithm. As we are trying to use linear features to depict the object structure, Hough transformation is appropriate for the task. Also combination of these is possible, here we have used edge linking algorithm and applied Hough transformation afterwards.

If we consider the 2D projection of three dimensional features, a space line which is also a line in 2D, a circle which is an ellipse in 2D and an ellipse which projection is an ellipse are best features to use respect to automatic feature classification with Hough transformation. These features are also good if we consider the definition of initial values for LSQ-estimation. Another curve like a b-spline is an ideal feature to depict many natural curves, but the problem of parameterisation and getting reasonable initial values restrict the use of it.

The Hough transformation is a widely known method for finding points which belong to a curve with a predetermined feature type. The disadvantage of the method has been its great computer consumption in time and memory. Many variations have been designed for decreasing the computational complexity of the method as well as improving the accuracy of the algorithm. The algorithms have been divided into standard algorithms and stochastic Hough transformations. One approach is to use statistical analysis to determine correct parameters in transformation space. This approach has been successfully investigated by J. Kittler, J. Illingworth, J. Princen and H.K. Yuen[5,6,7]. The disadvantage of this kind of approach is the complexity of the computation. The accuracy of the method as well as the reliability are remarkable. Another fascinating algorithm, proposed by L. Xu and E. Oja[8], is the *Random Hough Transformation* which fulfills both the expectation of accuracy and saving of storage space.

3.1 RHT Randomized Hough transformation

The Randomized Hough transformation RHT belongs to the category of probabilistic Hough transformations. The idea of the method comes from neural computing. By inspirations of Kohonen map, the algorithm uses converging mapping to determine the Hough parameters. The whole method relies on *random sampling, converging mapping* and *stepwise implementation of accumulation*. Xu and Oja have introduced the algorithm for usage with different kinds of score storage structures, but the dynamic list structure appears to be the best at the point of view of storage space.

The procedure goes by alternating the converging mapping and accumulation periods. The procedure requires predetermined values to terminate the process. The number of maximum trials for verifying a feature existence must be heuristically set. This of course depends on the case. With suitable values the algorithm finds quite accurately all points belonging to each feature and is much faster than other comparable methods. The processing scheme of RHT-algorithm is depicted below.

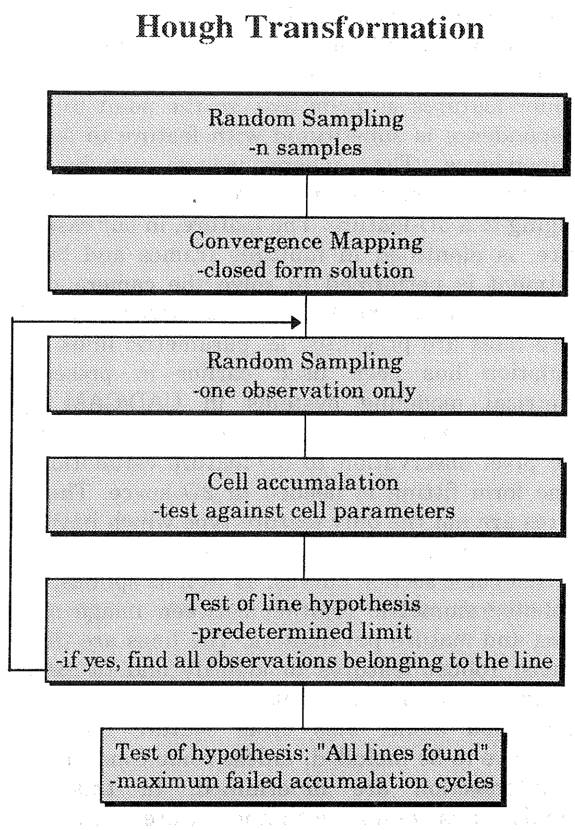

Figure 1. Procedure scheme of RHT.

The algorithm was applied for edges extracted with Canny operator, Figure 2. The edges falling into subpixel gap in Hough space are depicted in Figure 3.

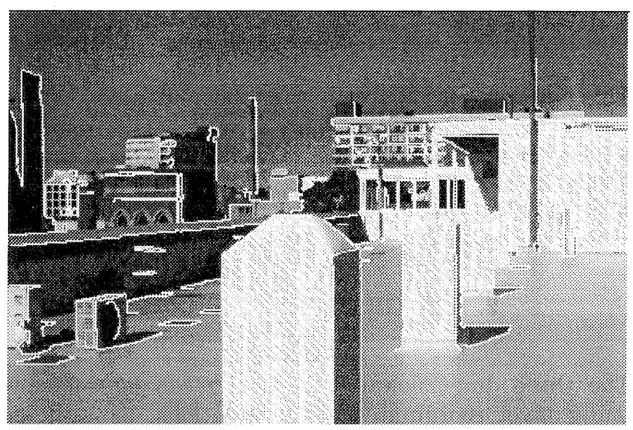

Figure 2. Results of the edge detection. Canny operator .

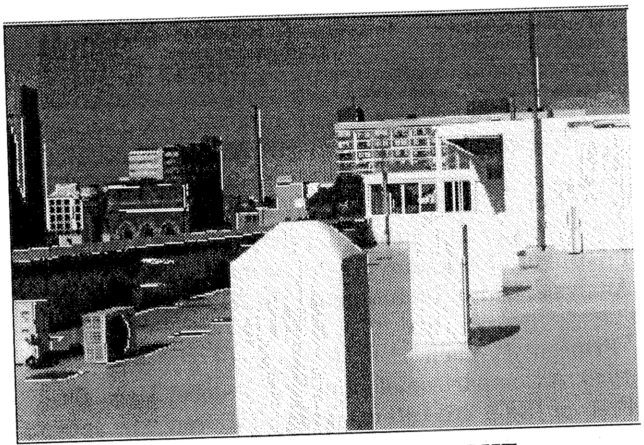

Figure 3. Lines found after RHT.

4. FEATURE MATCHING

In this chapter, the term *feature* means a 2D object in image space, which has characteristic properties belonging to the object, notation deriving from pattern recognition field. And this should not be confused with 3D linear features mentioned earlier. The feature matching then simply means solving the correspondence problem between *features* from different frames. When measuring images of a video sequence, the displacement of the current *feature* between consecutive frames cannot be too big. For this reason Hough parameters are good *feature* descriptors of a 2D image *feature*, both the length of the arc as well as the average strength of edges are suitable for the task. In case of line, the spatial coordinates of starting and ending point of the line are distinguished descriptors.

Matching of features between consecutive frames can be a ambiguous process. The first stage is to construct all combinations of feature pairs and to calculate similarity measures between them. The correlation between feature vectors presents one good measure. Often some kind of normalization is needed for correlation coefficients. Based on these similarity measures weights for each feature pair are determined.

Finding the correct feature pairs for the features of the first frame from the second frame, i.e. feature matching, can be done in many, different ways. One method widely used in any sort of situations is the *probabilistic relaxation*. The idea of this method is that the nodes near proximity effect on weights of the node. Relaxation is then an iterative process. The result can although depend on the order in which the nodes are updated.

A problem occurs when occluding particles appear. In a such case, some heuristic threshold value has to be set for a similarity measure to eliminate the affect. Also geometrical constraints like epipolar constraint can stabilize the matching. The assumption is that camera movement is smooth between the frames. This may restrict the search space and make the matching more robust.

5. FEATURE MODELING

Three dimensional form fitting can be done by using two dimensional image observations from two or multiple images, whose pose differ from each other for solving three dimensional parameters of the features. That means linear 3D features like lines, circles, ellipses, parabolas, hyperbolas, and b-splines are used instead of points to reconstruct the object. In photogrammetry D. Mulawa[2] presented the idea in his dissertation thesis in. There he used this kind of three dimensional parametric form of the features to depict shape and size of the objects. The idea of doing form fitting in three dimensional space means that no subpixel line detection is needed in image space. The whole estimation can be done in three dimensional space using original pixel observations.

The parametric presentation is very compact. The general form of curve can be presented as a set of points. In a case of a three dimensional curve its trace consists a certain set of points P_i.

$$\mathbf{F_i} = \{ \mathbf{P_i} \} \tag{1}$$

In the parametric formulation we can find a common set of parameters u_i on which all points of curve are dependent. The general formulation of parametric presentation can be given as,

$$\mathbf{P_i} = \mathbf{P(u_i)} = \begin{bmatrix} \mathbf{x(u_i)} \\ \mathbf{y(u_i)} \\ \mathbf{z(u_i)} \end{bmatrix}$$

$$u_i = set\ of\ parameters\ of\ feature\ i \tag{2}$$

All parametric presentations are not unique without involving some constrains. For modeling purpose also constrains between features are possible. Constrains e.g. intersection of lines in three dimensional space, parallelism of lines etc., set by the operator can simplify and stabilize the estimation in the object reconstruction part.

To have a direct relation between image observation and three dimensional feature parameters gives us lot of redundancy in the estimation. We can have as many observation as edge points detected to determine the parameter values. The number of parameters is always small compared to number of observations we can have. And specially in our case with multiple video frames we can have massive number of observations connected to a single feature.

All photogrammetric basic tasks like exterior, relative and absolute orientation as well as space intersection can be resolved with feature based methods. But like in pointwise approach also in feature based procedure, the distribution of control features is essential for reliable results[9].

In modeling from video frames we use feature based triangulation to solve the modeling problem. After having a successful feature matching, the operator can point out control features for the system. Also similar kind of feature matching as in Chapter 4 can be applied for detecting control feature and 2D feature in image space correspondence. Before the final bundle adjustment, all those feature lines have to be rejected from sample space which are parallel or nearly parallel with the image base line. An exception is when those lines have a predetermined intersection between lines with different direction properties. Otherwise, this would lead into a singularity problem in the estimation. Also the use of parallel imaging strips like in aerial photography will eliminate this problem.

6. SUMMARY

The proposed algorithm combines automatic observation extraction, robust feature matching and use of linear features in modeling. The idea has been to utilize the power of LSQ-estimation by increasing the redundacy . This has been done in two ways; using massive number of frames in observation extraction phase and including all edge points of an object feature to determine its parameters.

The automation has not been carried out through the whole process, but only on most computing consuming phases. The operator inspection and assistance will be used in the process to guarantee the convergence of the system.

7. REFERENCES

[1] Mulawa, D.C., Mikhail, E.M., 1988. Photogrammetric Treatment of Linear Features. In: Int. Arch. Photogrammetry and Remote Sensing, Kyoto, Japan, Commission III.

[2] Mulawa, D.C., 1989. Estimation and photogrammetric treatment of linear features. UMI Dissertation information service, Purdue, p.312.

[3] Rosenfeld, A., Kak, A., 1982. Digital picture processing. Vol. 1-2, Computer Science and Applied Mathematics, Academic Press, Orlando, 2nd edition.

[4] Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Trans. on PAMI. Vol. PAMI-8, no. 6, pp. 679-698.

[5] Illingworth, J., Kittler, J., 1987. The Adaptive Hough Transform. IEEE Trans. on PAMI. Vol. PAMI-9, no. 5, pp. 690-698.

[6] Princen, J., Yuen, H.K., Illingworth, J., Kittler, J., 1989. Properties of the Adaptive Hough Transform. Proc. of 6th Scandinavian conference on Image Analysis, Oulu Finland, pp. 613-620.

[7] Princen, J., Yuen, H.K., Illingworth, J., Kittler, J., 1989. A Comparison of Hough Transform Methods. Proc. of IEEE 3rd International Conference on Image Processing and its Applications, University of Warwick, pp. 73-77.

[8] Xu, L., Oja, E., 1993. Randomized Hough Transform (RHT): Basic mechanisms, algorithms, and computational complexities. CVGIP: Image understanding, vol. 57, no. 2, pp 131-154.

[9] Heikkinen, J., 1994. Linear Feature Based Approach to Map Revision. Int. Arch. of Photogrammetry and Remote Sensing. Athens, Georgia, U.S.A. Commission IV Symposium, pp. 344-351.