Egomotion Estimation of a Multicamera System through Line Correspondences

Federico Pedersini, Augusto Sarti, Stefano Tubaro
Dipartimento di Elettronica e Informazione (DEI), Politecnico di Milano
Piazza L. Da Vinci 32, 20133 Milano, Italy
Telephone: +39-2-2399-3647, Facsimile: +39-2-2399-3413
E-mail: pedersin/sarti/tubaro@elet.polimi.it
http://www-dsp.elet.polimi.it/ispg/frame.htm

Abstract

In this paper we propose a method for estimating the egomotion of a calibrated multi-camera system from an analysis of the luminance edges. The method works entirely in the 3D space as all edges of each one set of views are previously localized, matched and back-projected onto the object space. In fact, it searches for the rigid motion that best merges the sets of 3D contours extracted from each one of the multi-views. The method uses both straight and curved 3D contours.

1. Introduction

In the past few years, close-range photogrammetry has witnessed a proliferation of methods for the automatic 3D reconstruction of objects from multiple CCD camera images. Among the numerous approaches available today, those that are based on stereo matching seem to be particularly promising. Such methods, however, can usually provide a reconstruction of just a portion of the scene surfaces, while it would be desirable to reconstruct the surfaces of the whole scene. As a matter of fact, automatic 3D reconstruction systems based on stereo-matching can only reconstruct the *visible* portion of surface. Such systems, in fact, typically provide a description of just the *front side* of the imaged scene or, when the surface is too large to fit simultaneously in all views, of just a limited portion of it. In conclusion, in order to obtain a complete scene reconstruction through stereometry, it is necessary to observe the scene from several significant viewpoints and put together the final reconstruction like a *patchwork* of partial reconstructions.

In order to be able to merge 3D data coming from different reconstructions, we need to accurately estimate the rigid motion that the acquisition system undergoes between two partial reconstructions. In order to do so, one could employ high-precision mechanical devices for positioning the camera system (or the object) before acquiring a multi-view. This *a-priori* solution of the ego-motion problem, however, is usually quite expensive and not very flexible. In alternative, one can perform detection and tracking of some image features throughout the acquisition process, and use the location of such features for estimating the camera motion. This last approach becomes particularly interesting when the features to be extracted are part of the scene to be reconstructed rather than being artificially added to it. Adding special *markers* to the imaged scene is, in fact, common practice in photogrammetry but, besides making the egomotion retrieval more invasive, it requires a certain expertise and slows down the acquisition process. Conversely, natural point-like features that are already present in the scene are difficult to safely extract and accurately locate. Scene features that can be quite safely detected are, instead, luminance edges. These features are more likely to be naturally present in the scene and rather easy to detect, which makes them good candidate features for egomotion estimation.

In this paper we present a method for estimating the egomotion of a multi-camera system from the analysis of 3D contours in the imaged scene. Being the method based on a calibrated multi-ocular camera system [3,4], the estimation is performed entirely in the 3D space. In fact, all edges of each one of the multi-views are previously localized, matched and back-projected onto the object space [5]. Roughly speaking, the method searches for the rigid motion that best merges the sets of 3D edges that are extracted from each one of the multiple views (see Fig. 1).

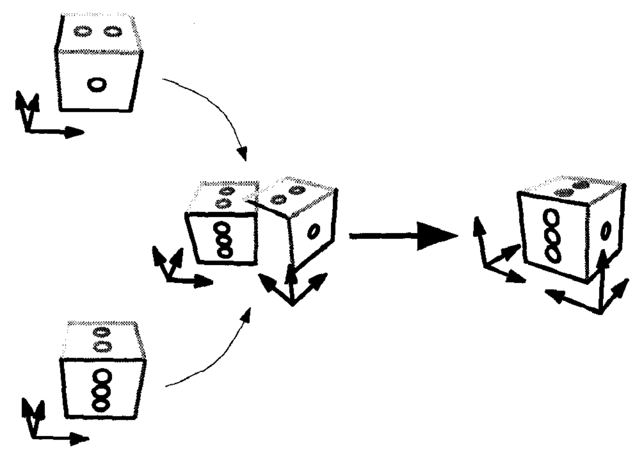

Fig. 1: egomotion estimation through line fusion: camera motion is determined as the one that best merges the 3D contours coming from different views multi-views

0-8186-8183-7/97 $10.00 © 1997 IEEE

2. Partial Reconstruction

The adopted camera model is basically a perspective projection onto an image plane, which is nonlinearly stretched in order to take the geometric distortion of the optics into account.

In order to obtain a complete 3D reconstruction of an object we need to merge a series of partial reconstructions, which can be obtained through a variety of techniques. In this paper we consider partial reconstruction from edge matching, so that we can use the same type of features for egomotion retrieval as well. In fact, partial reconstruction is based on 2D edge matching (stereo correspondence on the image planes), while motion estimation is based on 3D contour matching (edge correspondence in object space). It is important to emphasize the fact that, in order to be able to use edges for accurate egomotion estimation, we need them to be detected with great accuracy. We do this by first using a traditional edge detector, we then retrieve the subpixel location of the edge points through an interpolation process which takes the luminance gradient into account. Finally, a rule-based contour tracking method is employed for determining the correct connection between edge points.

The search for homologous edges on different views is performed along *epipolar lines*. Notice that using more than two cameras allows us to avoid problems of matching ambiguity. For example, with three cameras, not only can we always select the best pair of views for a specific stereo-correspondence (sharp intersection between edge and epipolar lines), but we can validate the matching through a check on the third view. In fact, the edge point must lie on the intersection of the two epipolar lines associated to the homologous edge points on the other views. Once the stereo correspondences are found, each set of corresponding contours is back-projected onto the 3D scene space by looking for the point at minimum distance from the three homologous visual rays.

3. Egomotion through Line Matching

The egomotion estimation method that we propose in this paper is organized in two mains steps. After having partitioned the available 3D contours in lines and curves, we proceed as follows:

1. rough egomotion estimation from straight contours:
    - matching of straight contours
    - motion estimation through minimization of the distance between homologous contours
2. egomotion refinement using curved contours:
    - matching of curved contours
    - motion estimation through a minimization of the distance between homologous curved contours.

Notice that, as a first approximation of the egomotion is already available, the matching of curved contours is a rather simple operation compared with the matching of straight lines.

3.1 Egomotion from Straight Lines

Line matching in 3D space is performed through a hypothesize-and-test type of procedure [2]. The first step of this method consists of formulating hypotheses on the possible couplings by selecting all those that do not violate some rules of congruence based on a set of geometrical constraints. By doing so, we drastically reduce the search space over which to test for matching correctness. At this point we can proceed with an exhaustive search through the above reduced set of hypotheses and select the match that maximizes an appropriate measurement of the matching quality.

Once the matching process is complete, the egomotion estimation can be performed rather easily by searching for the rigid motion that minimizes an appropriate *merging cost* function between two sets of 3D lines that pertain two different partial reconstructions. Notice 3D contours are generally reconstructed as chains of segments whose length and fragmentation may vary quite drastically from multi-view to multi-view. We thus proceed by first determining the 3D line portions that best fit (through linear regression) the chains of fragments of edges that have been recognized as straight. Then instead of measuring the distances between extremal points of two segments, we measure the distance between the extremal points of one segment and the line that the other segment lies upon (see Fig. 2). Such distances are used for defining the merging cost as follows

$$C_s = \sum_{i=1}^{N} \left[\left(d_i^b\right)^2 + \left(d_i^e\right)^2 \right].$$

In fact, the orientation of edges is usually less sensitive to fragmentation problems than their location in the 3D space [1,2].

3.2 Egomotion Refinment from Curved Contours

As already said above, curved contours are used for improving the accuracy of the egomotion's estimate. Although a matching process is required in this case too, this step is now simplified by the knowledge of a first

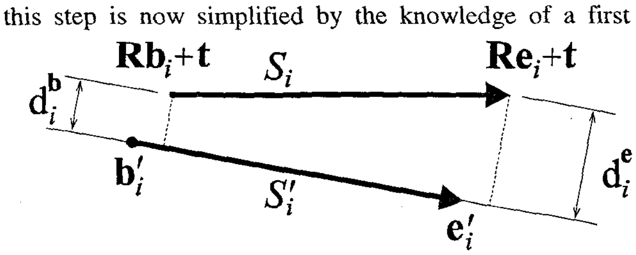

Fig. 2: evaluation of the merging cost of two straight 3D contours

approximation of the camera motion, determined from straight edges. In fact, by applying the pre-determined rigid motion to the set of curved edges, we can decide whether two curved edges are matched, depending on their global distance, which can be measured, with reference to Fig. 3, as

$$d_g = \frac{1}{2}\left(d(C,C') + d(C',C)\right)$$

where

$$d(C,C') = \frac{1}{N}\sum_i d(E_i, C')$$

$$d(E_i, C') = \left\| \overrightarrow{E_i E_i'} \right\|$$

The global cost function for motion refinement is of the form $C = C_s + kC_c$, where C_s and C_c are the merging costs associated to straight and curved contours, respectively, and k is weight for balancing the two contributes.

4. Examples of Application

The method has been extensively tested against convergence problems and has been applied to a series of trinocular acquisitions of real images in order to evaluate qualitatively and quantitatively the accuracy of the results and the speed of convergence. Furthermore, the performance of the proposed method has been compared with that of a previously studied method [6,7] based on point correspondences between artificially added markers. Quantitative results have been obtained by measuring the maximum thickness of the bundles of edges when superimposing different sets of them with the estimated motion parameters. The performance of the proposed method has been proven to be equal to or better than that of the point-based approach, resulting in a maximum bundle size of about 100 ppm in all tests (after merging all 3D edges coming from 20 multi-views).

In Figs. 3 and 4, results on 3D data merging are reported for two different objects in both cases of

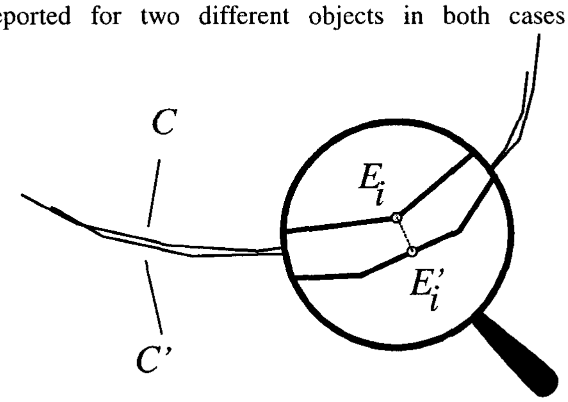

Fig. 3: 3D curve matching: evaluation of the distance between two polylines

egomotion estimated through point and line correspondences. In the first case the cost function is a rigidity constraint based on the distance between reconstructed 3D points of different 3D data sets. Such points are markers that have been artificially added to the scene (white dots placed on the object's support). In the second case the egomotion is computed with the method proposed in this paper. Even though no artificially added markers have been used for the estimation, the accuracy of the estimate is comparable with that obtained through point-matching.

5. Conclusions

In this paper we proposed a method for estimating the motion of a calibrated multiocular camera system from the multi-views of the scene to be reconstructed. The method is based on the analysis of luminance edges for performing both partial reconstruction and egomotion retrieval. Motion estimation is performed in such a way to best fuse the 3D data extracted from the available multi-views. In particular, straight 3D lines are used for determining a first approximation of the egomotion, which is then refined by using curved 3D contours. The method has been tested over a variety of real scenes, proving its performance comparable with what can be done with 3D point-matching on artificially added markers.

References

[1] Ayache, N., 1991. *Artificial vision for Mobile Robots.* MIT Press.

[2] Z. Zhang, O.D. Faugeras: "3D dynamic scene analysis: a stereo based approach", Springer-Verlag, 1992.

[3] Tsai, R.Y., 1987. "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using off-the-shelf TV Cameras and Lenses". *IEEE J. Robotics and Automation*, Vol. RA-3, No. 4, pp. 323-344.

[4] Pedersini, F., Tubaro, S., Rocca, F., 1993. "Camera Calibration and Error Analysis. An Application to Binocular and Trinocular Stereoscopic Systems". 4th Int. Workshop on *Time-varying Image Processing and Moving Object Recognition*, Florence, Italy.

[5] F. Pedersini, S. Tubaro: , "Accurate 3D reconstruction from trinocular views through integration of improved edge-matching and area-matching techniques." *VIII European Signal Processing Conference*, September 10-13, 1996, Trieste, Italy

[6] F. Pedersini, A. Sarti, S. Tubaro: "3D Motion Estimation of a Trinocular System for a *Full*-3D Object Reconstruction". *IEEE Intern. Conf. on Image*

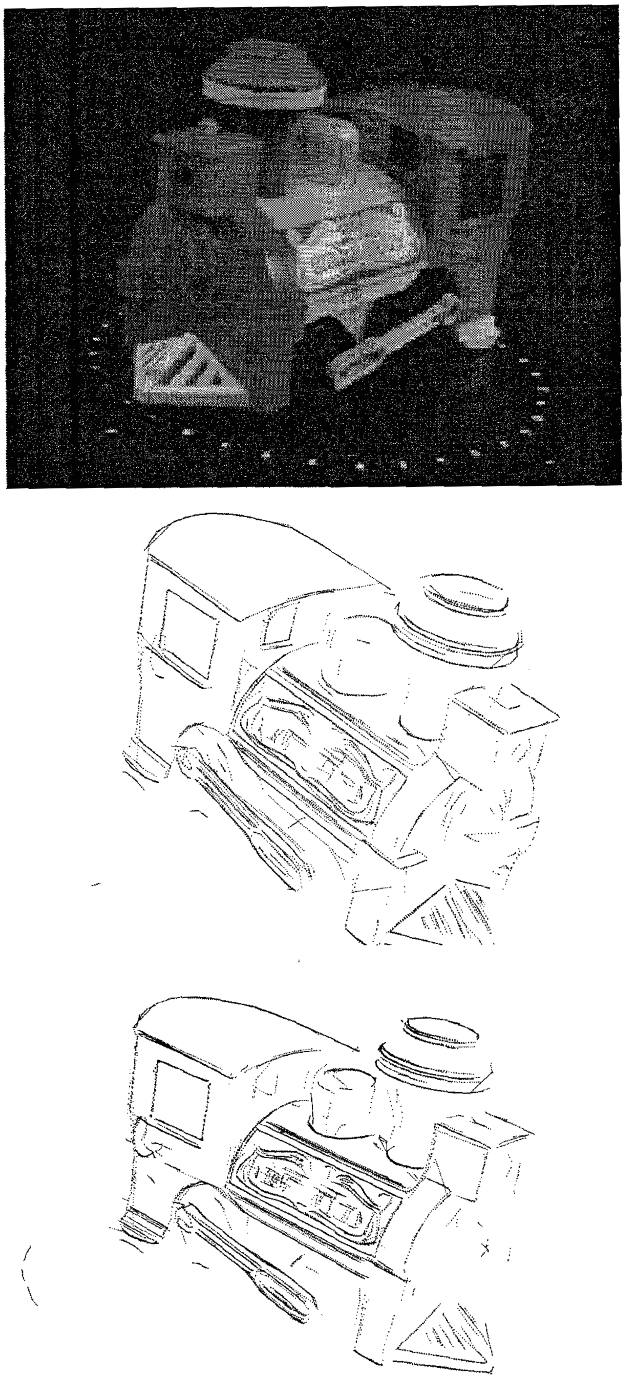

Fig. 4: From top to bottom: view of the original object; fusion of all 3D edge sets through point correspondence; fusion of all 3D edge sets through 3D contour matching.

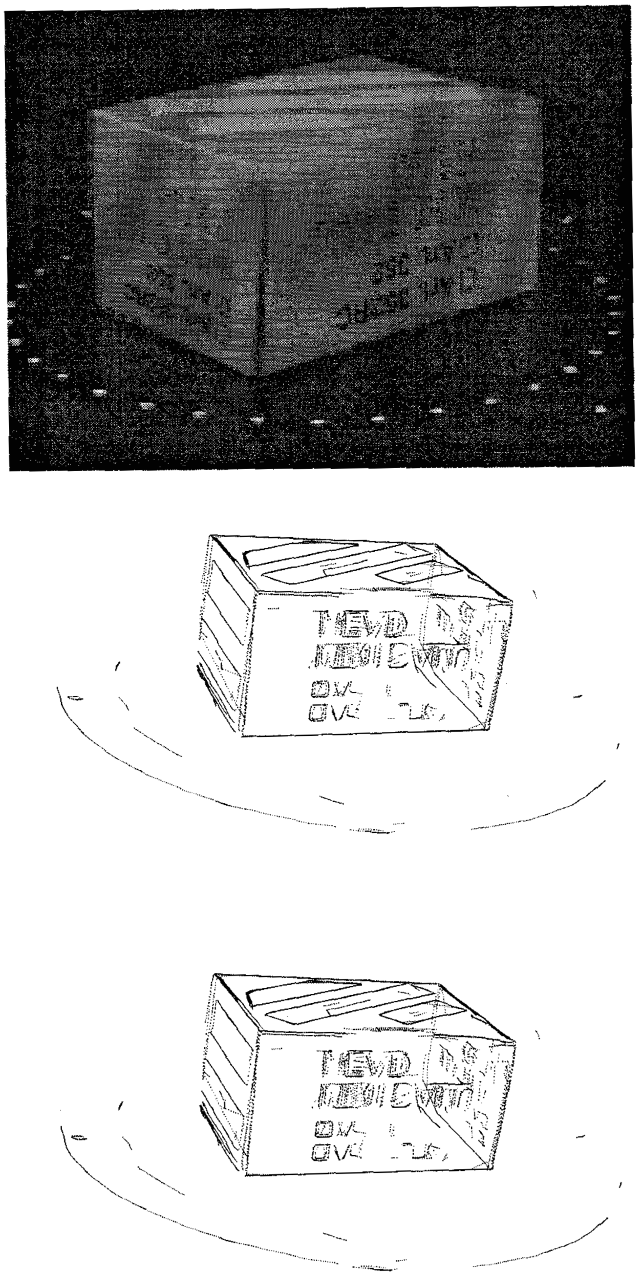

Fig. 5: From top to bottom: view of the original object; fusion of all 3D edge sets through point correspondence; fusion of all 3D edge sets through 3D contour matching.

Processing, September 16-19, 1996, Lausanne, Switzerland.

[7] F. Pedersini, A. Sarti, S. Tubaro: "A Multi-view Trinocular System for Automatic 3D Object Modeling and Rendering." *XVIII International Congress for Photogrammetry and Remote Sensing*, July 9-19, 1996, Vienna, Austria.