ACCURATE 3-D RECONSTRUCTION FROM TRINOCULAR VIEWS THROUGH INTEGRATION OF IMPROVED EDGE-MATCHING AND AREA-MATCHING TECHNIQUES

Federico Pedersini, Stefano Tubaro
Dipartimento di Elettronica e Informazione (DEI), Politecnico di Milano
Piazza L. Da Vinci 32, 20133 Milano, Italy
Tel: +39-2-2399-3647, Fax: +39-2-2399-3413
e-mail: pedersin/tubaro@elet.polimi.it

ABSTRACT

This paper describes a method for obtaining a reliable 3D reconstruction of close-range objects by properly combining edge- and area-based matching techniques. The adopted acquisition system is a set of three calibrated low-cost CCD cameras. By using an accurate camera model and camera calibration, the method is capable of working with any camera setup. The proposed technique has been tested on some real scenes with encouraging results. Some of these experimental results are presented here.

INTRODUCTION

Measurement and reconstruction of close-range object surfaces is more and more often done in an automatic way by means of multi-camera systems, through detection, matching and back-projection of image features.
A class of image features that is very often used for stereo matching is that of *luminance edges*. Image edges are particularly suitable for 3-D reconstruction because of the intrinsic precision and reliability of their localization on the image [3]. Luminance edges, however, can only provide us with sparse sets of 3-D points, as they are concentrated where the object surface is jagged or highly textured. For this reason, a different approach, in which the luminance profile of small *image patches* is used as image feature (*area-based* approach), can be usefully employed in correspondence to image regions where the luminance profiles are smooth. This approach, however, suffers from perspective and radiometric distortion between the different camera viewpoints.
In this paper an area-based matching technique which takes into account for both the above distortions is proposed and described. Through this technique it is possible to estimate both 3-D position and local surface orientation of a scene detail, starting from the luminance patch that corresponds to its projection in one of the images. As this estimation is an intrinsically ill-conditioned operation, it is necessary to provide the algorithm with an accurate “first guess” of the 3-D position to be determined. Hence, the proposed reconstruction algorithm works as follows: first edge-based reconstruction is performed and the 3-D coordinates of luminance edges are determined; then, a first reconstruction of the 3-D scene shape is obtained through an appropriate 3-D surface interpolation, starting from the 3-D edges. Surface interpolation is performed by means of a modified version of the *thin-plate spline* algorithm (D.S.I. [2]). This choice allows us to cope with depth discontinuities, that always occur at object borders. The surface we obtain can thus be used as "first guess" for area-based reconstruction. The shape of the interpolated surface will, in fact, be refined through the insertion of new 3-D points.
The paper is organized as follows: in the next Section, a brief description of the adopted camera model is reported. After that, Sections 3 and 4 are devoted to the description of the edge- and area-based reconstruction techniques. The performance of the proposed multi-approach algorithm has been evaluated by employing it in the analysis of several real scenes, acquired with a three-camera system. Some of the obtained results are described in Section 5.

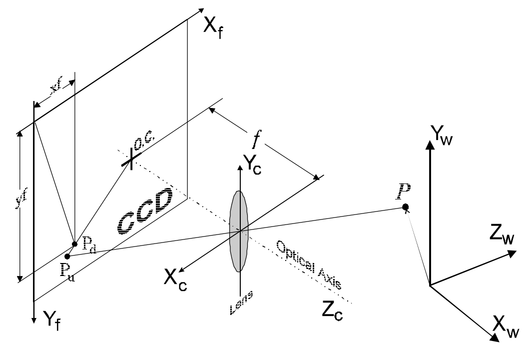

Fig. 1. The adopted camera model.

THE CAMERA MODEL

What it is normally meant with *camera model* consists of the set of mathematical relationships that relate the position of a point in the three-dimensional space imaged by the camera to the projection of that point on the image plane. The scheme of the adopted camera model is shown in Fig. 1, where three reference frames are visible: the *world reference frame*, attached to the imaged scene; the *camera reference frame*, attached to the camera, and the *image reference frame*, where the axes are attached to the digital image.
The equations of the camera model define a mapping of a generic point P of world-coordinates $P_W=(X_W, Y_W, Z_W)$ into

the 2D coordinates (x_f, y_f) of its projection on the image plane. These equations are:

a) Change of reference frame, from *world*- to *camera*-coordinates

$$P_{cam} = \mathbf{R} \cdot P_W + \mathbf{T};$$

R being the rotation matrix and **T** the translation vector.

b) Perspective projection of a scene point to the image plane (the center of projection is the center of the lens and the projection plane is the camera CCD sensor):

$$p_u = P_{cam} \cdot \frac{f}{z_{cam}}$$

c) Lens distortion shift of the image point p_u, predicted via perspective projection, to the actual position p_d. When standard-resolution CCD cameras are being used, only the radial component of distortion is normally considered. Radial distortion is usually approximated by a power series:

$$r_u = r_d \cdot \left(1 + k_3 \cdot r_d^2 + k_5 \cdot r_d^4 + \ldots\right).$$

This series can be generally truncated at the fifth order (only the first two coefficients are used) as the residual error results as being far below 1 pixel [1].

d) Change of coordinate frame from camera-coordinates $p_d=(x_d,y_d)$. to image-coordinates $p_f=(x_f,y_f)$.

$$x_f = C_X + \frac{x_d}{d_X}; \qquad y_f = C_Y + \frac{y_d}{d_Y};$$

d_x and d_y being the horizontal and vertical size of an image pixel, respectively, and (C_x,C_y) the *image*-coordinates of the optical center *OC*, which is the intersection between the optical axis and the image plane. As we can see from the above description, the camera model is completely specified by a set of *camera parameters*. All these parameters are estimated through an appropriate calibration procedure, that we have developed as an improvement of Tsai's algorithm [1]. More precisely, Tsai's algorithm is first applied to a simpler camera model, in order to provide a first estimation of most of the parameters. A refinement of all parameters is then carried out, through a nonlinear parametric estimation procedure based on the minimization of the error between the actual image position of a set of known reference points and the position predicted by the estimated model [1,5].

THE EDGE-BASED APPROACH

After the acquisition of an image triplet of the considered object, all significant edges are detected by means of a modified version of Canny's method [4]. After detection, edge selection is performed on each image, in order to preserve only those edges that are likely to correspond to significant features of the scene objects. The *relevance* of each edge is, in fact, quantified on the basis of its length (a minimal length is set as a threshold) and its shape. The selected edges are then labeled and divided into chains of small segments. For each one of these segments, the stereo-corresponding edges on the other two images are searched for through an appropriate procedure.

Notice that, because of the lens distortion, which is included in the camera model, the epipolar lines are no longer straight. Consequently, for each edge point we need to determine the "distorted" epipolar line on each one of the other two images. The search for an homologous edge is thus performed in the *epipolar space*, represented by these curves. Using three cameras allows us to select the best pair of views for a specific edge stereo-correspondence and validate it through a check on the third view. Congruence rules on the 3-D back-projections of each element are used for reducing the risk of errors and solving for possible ambiguities.

The output of the search algorithm (which is applied starting from each one of the 3 images) is a set of matched triplets of edge points, each one belonging to one image. This information is analyzed by an *edge matching* algorithm, whose aim is to find the most likely set of matched triplets of edges. In fact, due to a different edge fragmentation in the three views, the algorithm must be able to find correspondences between triplets of *multiple edges*, as an edge in one image could correspond to several edges in another one.

As final step, each triplet of edges is back-projected onto the 3-D scene space. The accuracy of the backprojection also benefits from the availability of a triplet of images as each 3-D point is given by the point having minimum distance from the visual rays of each perspective view.

SURFACE INTERPOLATION

The above-described algorithm, like any edge-based approach, is usually capable of providing a 3-D set of points of the scene that is typically quite accurate, but very sparse. This fact may be mainly attributed to the fact that only small portions of the scene surfaces usually contain edges that are suitable for matching. In order to obtain a denser depth map, it is thus necessary to interpolate the depth information all over the scene. This could be done by interpolating the depth map as a 2-D field, in each one of the image planes [6]. Representing 3-D shapes as a 2D perspective map, however, causes inconsistency problems, particularly for camera geometry with non-parallel optical axes. Alternatively, we may generate directly a 3-D surface which passes through all the computed 3-D edges. Such a surface would be an approximation of the shape of the observed scene.

The depth function of a 3-D scene is characterized by discontinuities at object boundaries and discontinuities in its first derivative at object ridges. Inside an object surface, the depth function is continuous and quite regular. We thus need the interpolation algorithm to be able to generate smooth surfaces anywhere except some special locations. Such an interpolator, introduced by Mallet [2], it is known

as *Discrete Smooth Interpolation* (DSI), and is based on a modification of the *thin-plate spline* algorithm. Its capability of preserving discontinuities is obtained through the specification of both local and global surface roughness parameters, which account for the presence of discontinuities in the neighborhood.
An optimized version of this interpolator has been implemented and employed for the construction of a surface that approximates the shape of the scene, from the sole information coming from the 3-D edges. This surface can be used as a starting point for the area-based shape refinement method.

THE AREA-BASED APPROACH

Once the first 3D reconstruction is available, we can proceed with its refinement through area-matching,. This operation adds information in correspondence to smooth image regions, where there are no edges to match.
The luminance patches used by most area-matching techniques are normally assumed to have the same shape in all views [6]. It is quite clear, however, that this hypothesis is acceptable only whenever the angles between the viewing directions of the three cameras are not too large, which is not our case. In order to be able to use area-matching with strongly convergent cameras, we have developed a generalized area-matching technique which takes into account perspective distortion of the patch shape.
As a first step, a small region (reference patch) of one image is back-projected onto the 3-D surface approximating the scene. The 3-D surface is locally approximated by a plane (see Fig. 2). The resulting 3D patch is then reprojected onto the image planes of the other two cameras, by taking into account projective and radiometric distortion. The minimum of a similarity function between synthetic re-projected patches and corresponding patches of the actual images is searched for as a function of the position and the local orientation of the 3-D plane of the patch. As already said before, the employed search technique is dramatically speeded up by the fact that we already have an initial 3-D reconstruction of the object surface which greatly limits the search space.
A significant improvement in the performance of this method has been obtained by introducing two extra parameters in the re-projection of the luminance patches: a *gain* and an *offset* factor. By doing so we take into account the fact that the electrical characteristics of low-cost cameras are not exactly the same and, at the same time, we account for non-Lambertian behavior in the reflectivity of surface patches (specular reflections). A single offset factor is determined for each camera, while a different gain is determined for each luminance patch during the matching process.
If a reference patch produces reliable 3D information, it can be used for refining the 3D surface reconstruction (see Fig. 4). Once all reference regions have been considered, a new surface interpolation is carried out (see Fig. 5), with improvements in areas corresponding to smooth surfaces.

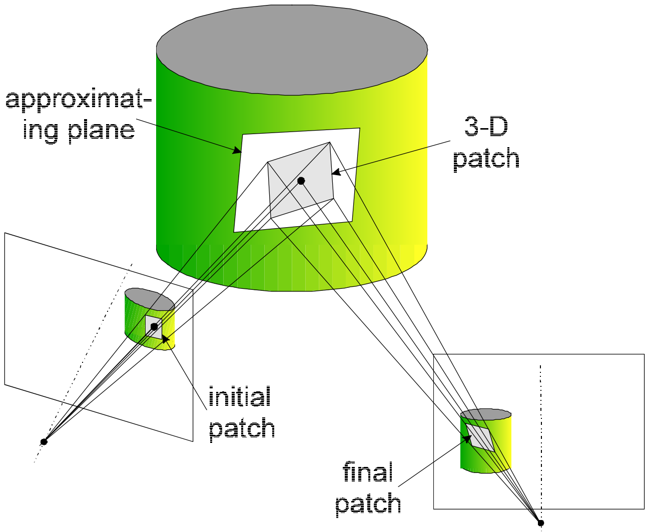

Fig. 2. Back-projection of an image patch.

EXPERIMENTAL RESULTS

Some experiments of 3-D scene reconstruction have been carried out on several test scenes. The results presented here concern the acquisition of the head of a dummy (fig. 3 shows one of the three views). In fig. 4 the global 3-D information extracted from the three views is shown; the dotted lines represent the extracted 3-D edges, while the scattered points represent 3-D information obtained by the area matching approach. Fig. 5 shows the reconstructed surface, generated by the DSI algorithm, which has been initialized with the 3-D edges and then refined with the 3D data obtained with the area matching procedure. This refinement leads to a significant improvement of the quality of the final reconstruction, especially on the areas corresponding to nose, chin and cheeks. In order to evaluate the reconstruction quality for the generation of virtual views, a synthetic view of the face of the dummy is shown in Fig. 6 after texture mapping. The virtual viewpoint is intentionally chosen to be very different from that of the cameras.

CONCLUSIONS

In this paper, a 3-D reconstruction system that combines edge- and area-matching techniques has been presented. Significant results have been obtained in the reconstruction of several real scenes. Some tests of 3D measurements accuracy have also been performed, considering simple and well-known objects. An accuracy of approximately one part per 10,000 has been reached.
Further research is being carried out in order to improve the 3D reconstruction by using more than one trinocular

view. Furthermore, "shape from shading" techniques are being developed in order to obtain 3D information from specular reflections of the objects and to synthesize more realistic virtual views.

REFERENCES:

[1] 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.

[2] Mallet, J.L., 1989. Discrete Smooth Interpolation. ACM Transactions on Graphics, vol.8, N.2, pp.121-144.

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

[4] Canny, J., 1986. A Computational Approach to Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.8, No.6, pp.679-698.

[5] 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.

[6] Otha, Y., Kanade, T., 1985 Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming, IEEE Trans. PAMI, Vol.7, N.2, pp.139-154.

Fig. 3. One of the original views of the dummy.

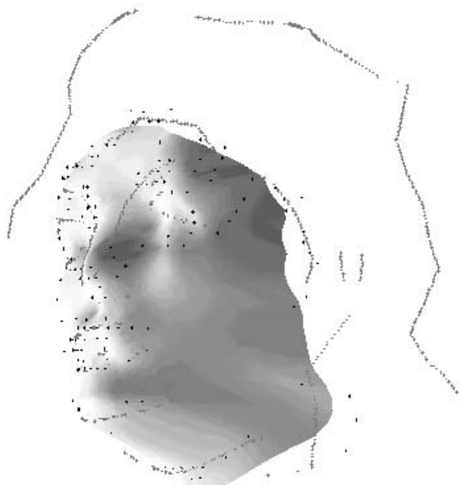

Fig. 5. Surface reconstruction from edge- and area-matching.

Fig. 4. 3-D features obtained with both edge- and area-based techniques.

Fig. 6. Virtual view of the head of the dummy.