MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

3—24 An Optimal Low Cost Solution For The 3D Free Form Object Recovering Pose Problem

Fakhreddine Ababsa, David Roussel, Malik Mallem
Laboratoire Systèmes Complexes, CNRS FRE 2494
Université d'Evry, FRANCE

Abstract

This paper present an efficient low cost solution for the 3D free form object recovery problem. We first describe the realized active vision system used to measure the 3D range image of the object, this system consists of a laser grid projector and a CCD camera. We present a robust sub-pixel method in order to detect image features. In the second part of this paper we present a 3D free form object recovery method, for that we elaborate a 3D/3D automatic matching technique based on invariant. We demonstrate the effectiveness and the optimality of our 3D recovery solution while implementing it with complex free form objects

1 Introduction

3D recovery of free form objects is an important research topic in robotics and computer vision. To successfully perform this task, 3D accurate measurements systems are needed. Triangulation with structured light is a well-established technique for acquiring 3D information on scene objects [1,2]. It projects a regular pattern of light onto the scene and hence creates artificial features on the surfaces of the objects that are easy to extract from the image.

In this paper, we first present a low cost and accurate 3D active vision system. It consists of a laser projector and a CCD camera. A regular grid of lines is projected onto the scene, intersect the surface of the objects and traces out a deformed grid. Such an encoded scene is then recorded by the camera. In our approach we consider the projected grid as a graph taking into account only nodes (lines grid intersection points) and the topological relationship between them. One of the major contribution of this work is the development of a robust sub-pixel model based approach [3] to extract grid nodes from captured images and also a solution for solving the correspondence problem between projected and original grid. The depth information is then retrieved by triangulation of registered features.

In the second part of this paper we present a 3D free form object recovery method which uses 3D measurements obtained by our vision system to determine the current position of the 3D object in the scene.

2 Vision System

2.1 Vision Workbench

Our vision workbench is constituted of a grid projector mounted on a two *dof* turret allowing its rotation according to two directions, in this way several regions of the scene's objects can be scanned. The projector projects a square grid of both five horizontal and vertical lines on the target object. Both the projector and the camera can translate along d_z direction which allow us to bring the vision system closer to the worktable and therefore allow us to scan smaller objects.

2.2 System Calibration

The coordinate systems used in calibration procedure are illustrated in figure (1). The developed calibration method is based on pin-hole optical models of the camera and projector.

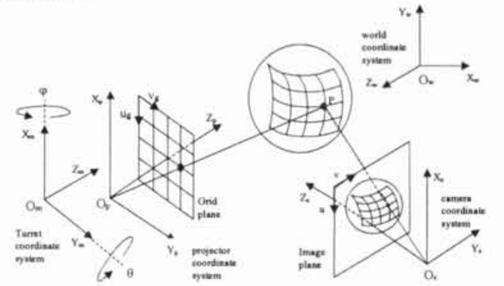


Fig.1: Coordinates system

2.2.1 Camera calibration:

Let *P(x,y,z)* be a 3D point in the world coordinate system (see figure 2), the relationship between P and its corresponding image point is expressed as [5]:

$$\begin{bmatrix} s\cdot u & s\cdot v & s \end{bmatrix}^T = M_c \cdot \begin{bmatrix} P & 1 \end{bmatrix}^T \qquad (1)$$

where s is a scale factor, M_c is a (3x4) calibration matrix.

Camera calibration involves estimating elements of the matrix M_c. By measuring enough known 3D points (we used 126 points) and their corresponding image points, M_c can be determined using least-square solution to linear equation (1). Practically, we use the grid projector to generate 3D.

2.2.2 Projector calibration

Our projector can rotate along two axes φ and θ in the rotating turret coordinate system. Angles φ and θ can be controlled with highest precision (0.01°), so we have developed a method which allow to calibrate the projector for any values of φ and θ.

First we calibrate the projector in its initial position ($\varphi=\theta=0$), the relationship between 3D grid node and its image grid point satisfies:

$$\begin{bmatrix} t\cdot u_g & t\cdot v_g & t \end{bmatrix}^T = M_p \cdot \begin{bmatrix} P & 1 \end{bmatrix}^T \qquad (2)$$

where t is a scale factor, M_p is a (3x4) calibration matrix.

Adress: 40, Rue du Pelvoux, CE1455-Courcouronnes 91020 Evry Cedex France.
E-mail: [ababsa|roussel|mallem]@iup.univ-evry.fr

As well as for the camera calibration, we generates several 3D points by the same method explained above. For each 3D point we know its corresponding grid node, we use least-square solution to linear equation (2) to estimate M_p elements. When the projector rotates along (O_mX_m) or (O_mY_m), only extrinsic parameters (i.e. the rotation and the translation between the world coordinate system and a projector coordinate system) change. So from M_p we extract intrinsic (i.e. projector internal parameters) and extrinsic matrices $I_p(3\text{x}4)$ and $A_p(4\text{x}4)$ respectively [6]. We can write then:

$$M_p = I_p \times A_p \tag{3}$$

and

$$A_p = \begin{pmatrix} R_{3\times3} & T_{3\times1} \\ 0_{1\times3} & 1 \end{pmatrix} \tag{4}$$

where R_{3x3} and T_{3x1} are the rotation matrix and the translation vector respectively from the world coordinate system to the projector one. When the turret rotates with angles φ and θ according to (O_mX_m) and (O_mY_m) axes respectively, The rotation matrix becomes:

$$R' = R_\varphi \times R_\theta \times R \tag{5}$$

In the same way, translation vector becomes:

$$T' = C_{m/w} \times \left(\left[R_\varphi \times R_\theta \right] \times O_{p/m} \right) \tag{6}$$

where $C_{m/w}$ expresses the transformation from the turret coordinate system to the world coordinate one. Equation (6) means that we calculate coordinates of the projector center first in the turret coordinate system, then we multiply them by $C_{m/w}$ matrix in order to obtain T' in world coordinate frame. Thus, the new extrinsic matrix A'_p expressed in world coordinate system can then be written as:

$$A'_p = \begin{pmatrix} R'_{3\times3} & T'_{3\times1} \\ 0_{1\times3} & 1 \end{pmatrix} \tag{7}$$

Finally, the global transformation matrix M'_p of the projector on its current position is given by:

$$M'_p = I_p \times \left(A'_p \right)^{-1} \tag{8}$$

This equation allows us to recalibrate the projector in any position without having to perform all the calibration procedure again.

2.3 Imaged Grid Extraction

In this section we will describe algorithms developed to extract the imaged grid with the highest precision and the ordering procedure. The imaged grid extraction procedure consists in two steps:

- Nodes detection.
- Grid reconstruction.

2.3.1 Nodes detection

Each node of the imaged grid is, locally, considered as a cross center. Thus, we propose a model based method to detect the cross center in a gray level image and which takes as a basis the optical response of target objects. The proposed luminance model allows the modeling of an any shape cross (right or oblique cross) and is given by the following equation:

$$L(x,y) = a + b \cdot \left(e^{c\cdot(x-d\cdot y-e)^2} + e^{-f\cdot(y-g\cdot x-h)^2} - e^{-c\cdot(x-d\cdot y-h)^2 - f\cdot(y-g\cdot x-h)^2} \right) \tag{9}$$

This model applies to images containing only one cross which is not the case here where we have the whole imaged grid, so we must first extract from the original image, small images that we will call "sub-image" containing only one cross at a time.

Thus, the accurate nodes detection includes the following steps:

- Approximate detection of grid nodes.
- For each node, we extract a sub-image centered on it.
- We determine the model parameters of equation (9) using a non linear optimization method in order to fit the sub-image luminance surface.
- We compute the exact node position in the original image from computed parameters of the model.

For each isolated cross sub-image $I(i,j)$, we use Levenberg-Marquardt method to determine optimal parameters of the model which minimize the following criterion:

$$c = \sum_{i=1}^{N_l} \sum_{j=1}^{N_c} \left[L(i,j) - I(i,j) \right]^2 \tag{10}$$

2.3.2 Grid reconstruction

The procedures described above allow us to obtain a list of nodes that we must order to reconstruct the projected grid. We suppose that ordering constraint is respected. Therefore detected nodes in the "projected grid image" will appear in the same order that nodes in "source grid". In practice this constraint is always verified because scene objects have piecewise smooth surfaces which undulate slowly with respect to the grid spacing.

To organize detected nodes, we first extract, from the skeleton image of projected grid, horizontal and vertical grid curves, point per point. Then we select only curves which contains grid nodes. We organize these imaged curve nodes from left to right and from top to bottom. We reconstruct the projected grid by the determination of intersection nodes between ordered horizontal and vertical node curves. Finally, projected grid reconstruction provides solution for the correspondence problem between projected grid and original one, which allow us to achieve 3D points reconstruction.

3 3D Recovery

The main objective of 3D recovering is to find the geometric transformation (namely translation + rotation) applied on a known object in the world coordinate system. This process is performed by matching features extracted from the object model with features extracted from the object wherever it may be within the world coordinate system. Our matching method is based on the following concept: Since a free form object has a well defined surface normal that is continuous almost everywhere except at vertices, edges and cusps [7], hence, we can consider some particular object regions that we call surface patches and we determine the distribution of angles between the patch surface normal and the normal vectors directly around it. Such an angles distribution is invariant with respect to rotation and translation. Each surface patch in the object has its own angle distribution which could be identified even though the object changes position. The architecture of our 3D recovery system is presented in figure (2), it is a Model-based architecture using range images.

3.1 Model Data Base

Objects used in this study are modeled with a triangular mesh achieved from scattered 3D points of the object measured with our system vision and expressed on the world coordinate system. Accordingly, the model object is defined by a pair of. We also compute for each surface patch gaussian curvature ($K=K_1*K_2$) and mean curvature ($H=(K_1+K_2)/2$) where K_1 and K_2 are the principal curvature. In practice, we use a discrete curvature [8] that applies directly to triangulated data.

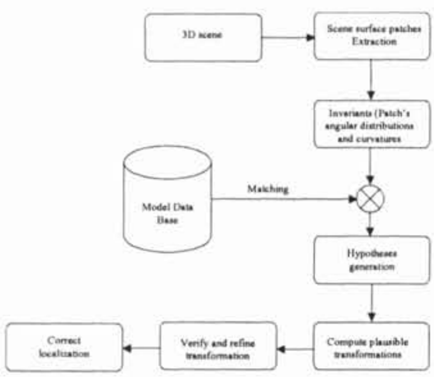


Fig.2: 3D recovery system architecture

3.2 3D/3D Matching and Verification

At recovering time, we use our vision system described in section (2) to extract several surface patches of the object, we choose to reconstruct local regions of the object of high curvature because in such regions angular distribution and curvatures are more discriminate and are relatively easy to identify in the hash table. For each reconstructed surface patch, we compute its angular distribution and curvatures, their values are compared to the ones stored in the model data base, if a correspondence occurs, a surface patch recognition hypothesis is generated. The verification module allow to discard all implausible recognition hypotheses and to only retain the good one. To achieve this, for each generated hypotheses we compute the corresponding geometric transformation, we back project the model features onto the image plane, and then retain the hypotheses for which back-projected features are the most close to their corresponding features [9]. The verification module needs to use the 3D localization one.

3.3 3D Localization

Let consider a matching hypotheses case generated by 3D/3D Matching module described above. Each correspondence between surface model patch and extracted one allow to identify a vertex and its normal when the object is in its new position. So 3D localization problem consists in finding the geometric transformation (translation T and rotation R) applied to the surface patches and which allow to displace them from their initial positions to their current positions with respect to the world coordinate system. Such a problem can be formulated as follow:

Let V_i be a 3D point of the object in its initial position (object model). $V_i^{'}$ be the point V_i displaced by the model transformation we are looking for, and let $\vec{N}_i(N_{xi} \quad N_{yi} \quad N_{zi})^T$ and $\vec{N}_i^{'}(N_{xi}^{'} \quad N_{yi}^{'} \quad N_{zi}^{'})^T$ the normal to V_i and $V_i^{'}$ respectively. $\vec{N}_i^{'}$ coordinates are given in world coordinate system by:

$$\vec{N}_{i/R_w}^{'}=R\times\vec{N}_{i/R_w} \tag{11}$$

In order to determine R, we have to minimize the following criterion

$$C=\sum_{i=1}^{N_m}\left\|\vec{N}_i^{'}-R\times\vec{N}_i\right\|^2 \tag{12}$$

where N_m denotes the number of matched surface patches. We use the rotation axis and angle representation to express the rotation matrix R. A least square method is used to compute an approximate solution for such a problem.

To determine the translation vector T, we notice that the $V_i^{'}$ points coordinates are given in the world coordinate system by:

$$V_{i/R_w}^{'}=R_{opt}\times V_{i/R_w}+T \tag{13}$$

however, the coordinates of the matched vertices V_i and $V_i^{'}$ in the world coordinate system are known. So for each two matched vertices, we compute the corresponding translation vector T_i as:

$$T_i=V_{i/R_w}^{'}-R_{opt}\times V_{i/R_w} \tag{14}$$

Finally, the optimal translation vector is given by:

$$T_{opt}=\frac{1}{N_m}\cdot\sum_{i=1}^{N_m}T_i \tag{15}$$

3.4 3D Refine Localization

When we apply the computed transformation to the model vertices and back project the transformed vertices onto the image plane we note an error between image features (in this case accurate nodes) and their corresponding imaged model features, this is due to the computation errors introduced by the data processing algorithms used in the several modules of our recovery. So to refine the computed transformation, we use the Levenberg-Marquardt method to minimize the mean square distance between accurate nodes extracted from scene image and their corresponding imaged model vertices.

Let D the criterion to minimize, it is defined as:

$$D=\sum_{i=1}^{N_m}\left\|\hat{m}_i-m_i\right\|^2 \tag{16}$$

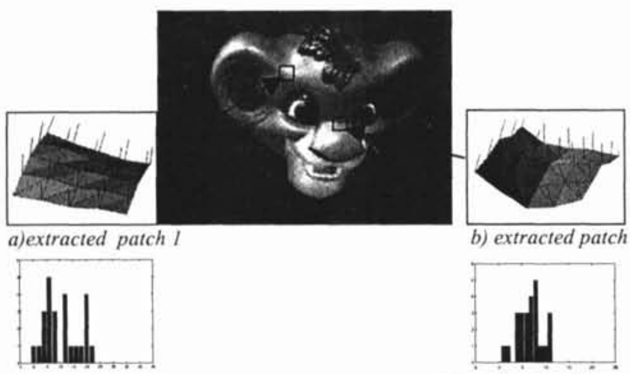
where $\hat{m}_i$ are the accurate nodes extracted from the object image and $m_i(u_i,v_i)$ are given by:

$$\begin{bmatrix} s.u_i \\ s.v_i \\ s \end{bmatrix} = Mc_{3x4} \cdot \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \cdot V_i \qquad (17)$$

where Mc_{3x4} is the global transformation defined by the camera calibration, R and T are the refine rotation and translation to determine. We use unit quaternion to define rotation R. A good initialization of the parameter vector P is necessary to assure the convergence of the algorithm. We use the transformation (R_{opt} , T_{opt}) obtained by the 3D localization module to initialize the vector P.

3.5 Experimental Results

The method presented in this paper to solve 3D recovery problem has been tested on a real free form object in this case a mask of a lion. So, we have first created the object data base, for that we have digitized the object using our vision system and meshed the obtained range image. We used at every time 25 vertices to constitute the several surface patches. Then, we have constructed a hash table as explained in subsection (3.1).



a)extracted patch 1 *b) extracted patch*

c) angular distribution (K>0 , H<0) *d) angular distribution (K>0 , H>0)*

Fig.3: extracted object patches

Fig.4: Object recovering

To perform localization of the object when it changes position, we scan at least two local regions of the object and we construct from extracted points the object surface patches that we try to recognize in the hash table. Figure (3), depicts the image of the object in its current position and in which regions to be extracted are marked by squares. Figures (3-a), (3-b), (3-c) and (3-d) illustrate the extracted surface patches and their angular distributions and surface curvatures. Using these features, our matching algorithm has well identify the extracted patches with model ones in hash table.

Once the 3D/3D matching is done, matched features are feeding to the 3D localization algorithm in order to find the current position of the object.

Figure (4) shows that when we apply the refine transformation to the model surface patches and re-project the obtained points onto the image, they well correspond to the extracted surface patches. This demonstrate the robustness and the pose accuracy of our method.

4 Conclusion

We have presented an efficient method to solve the 3D free form object recovery problem; To achieve our solution, we conceived a low cost vision system based on manufactured grid projector and a CCD camera. Calibration procedure of such a system has been also studied. We have developed a robust sub-pixel model based approach to extract grid nodes from captured images. Elaborated method is used in both camera calibration and 3D points reconstruction procedures. In the second part of this paper, we have presented in details the architecture of our recovery system. We described how, from 3D scattered points, we construct the object data base. We give also the mathematical solution of the localization and refine localization problems. The experimental results are very satisfying, the algorithm succeeds in determining the object pose by matching extracted surface patches with model ones using only their angular distributions and surface curvature. The obtained transformation is used to refine the object pose, this increases the robustness of our algorithm to noise and computation errors. The proposed method can be applied even though the object is partially occluded.

References

[1] G. Hu, G. Stockman, "3D Surface Solution Using Structured Light and Constraint Propagation", *IEEE trans. Pattern Analysis and Machine Intelligence*, Vol. 11 no. 4, April 1989, pp. 390-402.

[2] R.J. Valkenburg, A.M. McIvor, " Accurate 3D Measurement Using Structured Light System", *Image and Vision Computing*, Vol. 16 no. 2, February 1998, pp. 99-110

[3] R. Deriche, G. Giraudon, "a Computational Approach for Corner and Vertex Detection". *International Journal of Computer Vision*, Vol. 10, no 2, 1993, pp. 101-124.

[4] P.J. Besl, "The Free Form Surface Matching Problem". In H. Freeman, editor, *Machine Vision for Three Dimensional Scenes*, pages 25-71. Academic Press.

[5] A. Jonhson, M. Hebert. "Using spin-images for efficient object recognition in cluttered 3D scenes". *IEEE trans. Pattern Analysis and Machine Intelligence* , Vol. 21, no 5, pp. 433-449.

[6] O. Faugeras, "Three Dimensional Computer Vision: a Geometric Viewpoint". MIT Press, 1993.

[7] C. Dorai, A.K. Jain, "COSMOS- A Representation Scheme for 3D Free Form Objects". *IEEE trans. Pattern Analysis and Machine Intelligence*, Vol. 19 no. 10, October 1997, pp. 1115-1130.

[8] J.P Thirion, "The extremal mesh and the understanding of 3D surfaces". *International Journal of Computer Vision*, Vol. 19, no. 2, pp. 115-128.

[9] D.P. Huttenlocher, S. Ullman, "Recognizing solid objects by alignment with an image", *International Journal of Computer Vision*, Vol. 5, no. 2, pp. 195-212.