3D Shape Recovery and Registration Based on the Projection of Non-Coherent Structured Light

Roberto Rodella and Giovanna Sansoni
INFM and Dept. of Electronics for the Automation
University of Brescia
Via Branze 38, I-25123 Brescia - Italy
tel.: +39-30-3715446, fax.: +39-30-380014
e-mail: rodella@bsing.ing.unibs.it

Abstract
This paper describes the main features of an optical instrument for the acquisition and the registration of 3D images. The system exploits the projection of structured light, specifically developed to perform the measurement of free-form surfaces in an accurate way and over extended fields of view. The acquisition of multiple views is performed by controlling mechanically the rotation of the object under measurement: the corresponding rototranslation matrices are then used to align the partial clouds of points. Special attention is given to the calibration of the optical head, to enhance the flexibility of the measurement. The system layout, the overall elaboration procedure, and a number of significant experimental results are herein presented.

1 Introduction
3D shape recovery and modeling is the first step of the measurement chain in a number of applications, ranging from the reverse engineering process, to the dimensional quality control, to a number of machine vision and virtual reality applications [1]-[3]. The measurement requirements vary in dependence of the considered problem, and lead to different solutions, either on a contact or non-contact basis. The former approach is characterized by higher accuracy, and is generally time crunching, the latter allows a quick determination of very dense point clouds, with less accuracy, and in times strongly reduced with respect to contact probes.

In recent years a considerable amount of work has been carried out to increase the performances of 3D systems based on optical techniques, either passive or active. To this last category belong the methods which exploit the projection of light, either coherent or incoherent, to allow an easy solution of the so-called correspondence problem. Interferometry and moiré-based methods have been developed to get the depth information over small ranges at high resolution [4], time of light methods are successfully implemented over extended measurement ranges [5], while triangulation methods, based on scanning or fixed optical heads, cover the intermediate measuring interval [6]-[8].

In our laboratory, a 3D vision system based on the projection of structured light has been implemented. It exploits the projection of suitable light patterns, their acquisition by means of a video-camera and their elaboration to get the depth information. The work performed so far has been aimed at enhancing the flexibility of the system as well as its robustness, to the environmental light conditions, and to the possibility of implementing different projection and signal-demodulation approaches, to adapt to the measurement to a wide typology of objects. However, the system is able to perform the measurement of single views, the depth information is expressed with respect to a physical reference surface and within a coordinate system centered in the video-camera sensor plane. Despite the measurement performances were very appealing, as to the resolution of the measurement and the rather extended dimension of the illuminated area, (0.02% of the measuring range, over working areas of about 300 x 300 mm) the overall performances of the system were strongly limited by the inability to perform in an automatic fashion the multiview acquisition and registration, within a global reference system.

This limitation has been recently overcome by equipping the system with the tools for the automatic recovery and registration of multiple views. The use of a physical reference has been limited to the calibration phase of the optical head, and must be performed only when the optical head is reconfigured. During this
phase, a virtual reference is determined, with respect to which the depth of the surface illuminated by the projector and framed by the video-camera is computed. The object is placed on a rotation stage and the multiview acquisition is carried out in correspondence with the rotation of the object, mechanically controlled. Registration and alignment of different point clouds, usually realized by means of automatic software [9]-[10], is performed by using rototranslation matrices, determined by controlling the mechanical setup of the system. At the present stage of development, this procedure has to be thought as the initial step of a more complete one, involving the algorithmic derivation of pose and orientation of the object by means of 3D markers placed in the scene.

The characteristics of the measurement are the same as those observed for the gauging of single views; moreover, the rather large dimensions of each view allow to minimize the number of point clouds to be registered. Working volumes of about 300 mm x 300 mm x 200 mm (width x height x depth) can be completely covered by four views, with a precision of 0.1 mm.

In this paper, the main features of the sensor are presented. A significant set of experimental results is also reported.

2 Layout of the Technique

The basic layout of the system is presented in Figure 1. It exploits active stereo vision: an LCD projector projects suitable patterns of structured light on the target object, these are acquired, along a different direction with respect to that of projection, by a video camera. The deformation induced on the patterns by the object shape allows suitable light coding: this coding solves in an easy way the well-known correspondence problem.

The coding procedure combines the Gray Code method with the Phase Shift method [11]-[13]: it is based on the projection of a sequence of N=11 different black and white fringe patterns on the target object. The first seven patterns (n=0..6, in the time space) lead to a 7 bit Gray Code. The last four patterns (n=7..10) are obtained by shifting the eighth one of the Gray Code sequence by a quarter of its period in the space domain. Suitable binarization and elaboration of the images allow the determination of a matrix of codings: each one corresponds to a specific direction of light from the projector, and is represented by an 8-bits word where the first seven bits are boolean values (resulting from the binarization of the Gray Code pattern), the last one is a real value in the range 0 - 1 (resulting from the elaboration of the four shifted patterns). A one to one correspondence exists between pixels on the CCD and the matrix elements; thus, in the following both will be identified by coordinates (j,k).

The coordinates of the object points are obtained using suitable triangulation, exploiting a couple of matrices: the former, thereafter called Bit Plane Stack of the Object (BPSO) is determined in correspondence with the target under measurement, the latter, thereafter called Virtual Plane Matrix (VPM) is evaluated in correspondence with a plane surface, taken as a reference, during the calibration phase of the measurement. This procedure, to be performed only once after the setup of the system, is detailed in Section 3.

The use of a rotation stage allows the multiview acquisition: a data rototranslation procedure, involving precalculated pose matrices, performs the data alignment and allows the reconstruction of the whole object.

The geometry of the optical head can be varied in dependence of both the field of view and of the resolution required by the measurement. Moreover, the rotation stage can be moved with respect to the optical head to increase the flexibility of the measurement.

The overall measurement procedure is splitted in two steps: the former is aimed at the acquisition of a single view, the second to the multiview registration. In the following sections these phases are described.

3 Determination of the Virtual Plane Matrix (VPM)

The system configuration for the determination of VPM is sketched in Figure 2. Here, parameter d is the
system baseline, i.e.: the distance between the exit pupil P of the projection unit and the entrance pupil C of the acquisition unit, L is the distance of the baseline from plane R. The global coordinate system is placed in the video camera sensor plane (CSP) with origin in C, axes X- and Y- parallel respectively to the rows and columns of the CCD matrix and axis Z- coincident with the video camera optical axis. The discrete coordinates at CSP are j and k. The portion of plane R framed by the video camera is FW wide along X- and FH along Y-. The projection unit forms black and white fringes parallel to Y-: correspondingly, the light planes which can be identified in the 3D space are oriented as depicted in Figure 2, for two different light planes denoted respectively by LP1 and LP2.

The procedure to determine VPM is aimed at storing in each element $(j,k)$ of the matrix, the coding of the light direction from the projector that impinges plane R and is seen by the video camera at pixel $(j,k)$. This coding is a real value, expressed as:

$$I(j,k) = \hat{I}(j,k) + \frac{2}{\pi} \Phi(j,k)$$  \hspace{1cm} (1)

Eq. (1), highlights that $I(j,k)$ is obtained by adding to coding $\hat{I}(j,k)$, obtained by means of the Gray Code projection, an additional term, $\Phi(j,k)$ deriving from the Phase Shift approach. In fact, each pattern of the Gray Code sequence is acquired by the video camera and a suitable thresholding algorithm associates either the logic value ‘0’ or ‘1’ to the grey level of pixel $(j,k)$. Using seven patterns a 7 bits Gray Code is generated; it is then converted into integer value $\hat{I}(j,k)$ by a simple look-up table operation. The phase coding $\Phi(j,k)$ may be expressed as follows:

$$\Phi(j,k) = \Phi(j,k) + \Delta\Phi(j,k)$$  \hspace{1cm} (2)

In Eq. (2), $\Phi(j,k)$ is obtained as:

$$\Phi(j,k) = \tan^{-1} \left( \frac{I_1(j,k) - I_2(j,k)}{I_0(j,k) - I_2(j,k)} \right)$$  \hspace{1cm} (3)

In this equation, $I(j,k)$ is the intensity map of the $i^{th}$ shifted pattern $(i=0, 1, 2, 3)$, expressed as:

$$I_i(j,k) = A(j,k) + \frac{B(j,k)}{2} \cos \left[ \Phi(j,k) - \frac{\pi}{2} \right]$$  \hspace{1cm} (4)

where $A(j,k)$ and $B(j,k)$ stand for the average brightness and the fringe contrast respectively. $\Delta\Phi(j,k)$ represents a phase correction term experimentally evaluated [13].

The phase coding $\Phi(j,k)$ obtained in Eq. (3) gives a non-univocal description of the light ray seen by the video camera at pixel $(j,k)$, due to the fact that the range of non-ambiguity is limited to the interval $0-2\pi$. However, this limitation is overcome using the Gray Code contribution. Thus, both an extended measurement range and a good resolution are possible in this approach, while avoiding problems correlated to the unwrapping of the phase.

4 Procedure for the Acquisition of a Single View

During the acquisition procedure the measurement head is rotated by $90^\circ$ round the baseline: no physical
reference is necessary for the measurement of either single views or whole objects; instead, the range information is obtained by acquiring BPSO following the coding procedure above described, and by comparing it to VPM. Figure 3 represents the system geometry (the figure plane is perpendicular to Y-), and helps to understand the triangulation principle applied. For simplicity the determination of the coordinates of only point \( P_0 \) is shown, even though the method is valid for all the object points acquired by the video camera. In this figure a virtual plane is depicted: VPM is the corresponding matrix of codings. It is well evident that, in the presence of the object, the light coding seen by the video camera along direction AC is the same as that one corresponding to direction BC on the virtual plane. For this reason the difference between object matrix BPSO and virtual plane matrix VPM yields to a value proportional to distance AB. Distance \( Z_{P_0} = K'P_0 \) is evaluated by considering that triangles \( A\hat{P}OB \) and \( C\hat{P}OP \) are similar; thus, the following equation holds:

\[
Z_{P_0} = \frac{L \cdot d}{AB + d}
\]

The accurate determination of distance AB is not trivial. In fact, a simple difference of light codings does not adequately express AB, due to the stripe broadening along X-, determined by the crossed optical axes geometry of the system. A suitable calibration procedure using reference objects has been developed in order to compute with high precision the value of angle \( \alpha \) between PSP and the virtual plane. At the same time, the fine adjustment of values \( L \) and \( d \) from initial values, roughly measured by means of a ruler, is carried out. The detailed description of this calibration method is presented in [13].

Finally coordinates X- and Y- of \( P_0 \) are determinated using the following equations, which implement the 2D to 3D mapping with the correction of the error introduced by the parallax effect:

\[
X_{P_0} = \frac{FW}{2L}\left( \frac{2k}{K-1} \right) Z_{P_0} + \frac{FW}{2}
\]

\[
Y_{P_0} = \frac{FH}{2L}\left( \frac{2J}{J-1} \right) Z_{P_0} + \frac{FH}{2}
\]

In Eqs. (6) \( J \) and \( K \) are the resolution of the CCD matrix.

As an example of the system performance, the measurement of the wood-made sculpture in Figure 4 is shown. The field of view is about 400 x 300 mm along X- and Y- respectively, with a depth range of 150 mm. The object is hand-sculpt, and dark-green colored. The

---

Figure 4 Example of an object under measurement: wood sculpture.

Figure 5 Cloud of points (a) and syntetic shading rendering (b) of the wood sculpture.
measured frontal view of this target is presented in Figure 5, both as point cloud (5.a) and as synthetic shading rendering (5.b). The measurement is characterized by a resolution equal to 70\( \mu \)m. This results in the good visibility of the surface details due to the wood carving, as shown in Figure 6. Here, the face and the lateral part of the object are zoomed. The acquisition and elaboration time required to perform this measurement is about 20 sec, on a Pentium II 333. It is worth noting that high immunity to both reflectivity and color of the surface is guaranteed by suitable preprocessing of the images. The overall elaboration time is comprehensive of the time required by these procedures.

5 Procedure for the Multiview Acquisition and Registration

The acquisition of the overall object surface is performed by placing the object on the rotation stage and by rotating it in a suitable number of positions. The rotating structure is placed in front of the measurement head in such a way that the object surface occupies the central part of the working volume. This approach allows the maximization of the sensor performance. The procedure is based on the geometry described in Figure 7. Here two parameters besides \( L \), \( d \), and \( \alpha \) must be considered to account for the relative position of the rotation stage with respect to the optical head. These are AXISD (AXIS Distance) and OBLD (Object Base Line Distance). The former represents the distance, along X-, of the rotation axis (supposed parallel to the Y-) from the left boundary of the field framed by the video camera. The latter is the distance between the rotation axis and the baseline.

A suitable procedure has been designed to accurately estimate parameters AXISD and OBLD: it is performed by tracking the position of a few 3D features of a calibration object placed on the rotation stage and moved at a number of different values of the rotation angle.

5.1 Data Alignment

Data alignment is performed by means of the transformation of the data in the rotation stage coordinate system \((X_R, Y_R, Z_R)\). The system origin is placed at the center of the rotation plate. \(X_R\) and \(Z_R\) are oriented as depicted in Figure 7. The rotation axis \(Y_R\) is perpendicular the figure plane. Calling \( X = (X, Y, Z) \) and \( X_R = (X_R, Y_R, Z_R) \) the vectors representing the coordinates of a point in both the coordinate systems, we obtain:

\[
X_R = RT \cdot X
\]  

where \(RT\) is the rototranslation matrix, expressed as:

\[
RT = \begin{bmatrix}
-\cos \phi & 0 & \sin \phi \cdot \text{AXISD} - OBLD \cdot \sin \phi \\
0 & 1 & 0 \\
-\sin \phi & 0 & \cos \phi \cdot \text{AXISD} + OBLD \cdot \cos \phi \\
0 & 0 & 1
\end{bmatrix}
\]  

The elements of the matrix in Eq. 8 are function of the translation parameters AXISD and OBLD and of the rotation angle \( \phi \). This angle measures the
counterclockwise rotation around $Y_R$ of the rotating plate with respect to the $X_R$ system.

The performance of the procedure for multiview registration has been applied to the measurement of a handle. Four partial views at 0°, 90°, 180°, and 180° were sufficient to acquire the complete object surface. The point clouds corresponding to each single view are depicted in Figure 8. The whole object model is shown in Figure 9. The surfaces overlap very accurately, resulting in a total error at the junction of the surface boundaries less than 100 µm. The same approach to the measurement of the object in Figure 4 leads to the whole point cloud in Figure 10. In this measurement case four views have been acquired, and the measurement performance is the same as before. It is worth noting that the system guarantees good performance even on rather large fields of view and allows the minimization of the number of partial views needed to reconstruct the complete model. Typical values of the time required to obtain the complete model are within 2 minutes.

6 Conclusions

In this paper, the basic characteristics of a system for multiview reconstruction of 3D point clouds have been presented. The system implements the procedures for the light coding, the calibration of the optical head and the registration of the images. From the extensive set of measurement performed, good resolution as well as extended measurement range has been evidenced. Image registration is performed by means of rototranslation matrices, and by taking into account the mechanical constraint of the system. The performance of the measurement can be furtherly improved by compensating for the image distortion. Moreover the possibility of translating the optical head along X- and Y- will increase the typology of objects which can be measured. The technique intrinsically yields to the acquisition of both intensity and color information. Projecting white light or using the environment illumination the gray level or color information of the object surfaces ($I$ or $[R, G, B]$) can be acquired and added to the object model. These capabilities are under development, and will be the subject of a paper in preparation.

Acknowledgements

The authors gratefully acknowledge the contribution of Dr. Matteo Carocci for his careful reading of the manuscript.

7 References


