

Tele-3D — Developing a Handheld Scanner Using Structured Light Projection

João Filipe Ferreira Jorge Lobo Jorge Dias

Institute of Systems and Robotics, Coimbra Pole
Departamento de Engenharia Electrotécnica, Universidade Coimbra, Polo II

Abstract – Three-dimensional surface reconstruction using a handheld scanner is a process with great potential for use on different fields of research, commerce and industrial production. In this article we will describe the evolution of a project comprising the study and development of a system that implements the aforementioned process based on two-dimensional images. We will present our current work on the development of a fully portable, handheld system using cameras, projected structured light and attitude and positioning measuring sensors — the Tele-3D scanner.

I. INTRODUCTION

The problem of 3D spatial structure recovery and reconstruction has been the object of discussion in the past two decades for people involved in areas as diverse as the obvious computer, physical, or medical sciences, and as the perhaps less obvious, but nevertheless important, subjects as documentary and film production, down to all industries in general [1].

This article presents the study and development of a 3D information recovery and reconstruction system (commonly known as *3D scanner* or *3D digitiser* – although the latter is more encompassing and thus more correct, the first is more common and, for that reason, will be the designation to be used from now on), the *Tele-3D handheld 3D scanning system*. This project is an extension of our previous work (fully described in [1]), which concerned the development of a traditional, fixed laser 3D scanning system.

We will present, consecutively, the theoretical background for our project, the system's architecture and the implementation issues we are currently addressing.

II. THEORETICAL BACKGROUND

In the following subsections we will introduce the system's mathematical, geometrical and physical background.

A. The Projected Structured Light 3D Scanner

The first step for the recovery and reconstruction of three-dimensional scenes from two-dimensional images is to determine how to extract the required three-dimensional data from the two-dimensional information available on those images.

The geometry involving a generic structured light 3D scanning system is presented in figure 1 [1].

This work is supported by the Tele-3D Project, CERN/FCT. Corresponding authors: João Filipe Ferreira, Jorge Dias, E-mail: {jfilipe,jorge}@isr.uc.pt, URL: www.isr.uc.pt/~jorge.

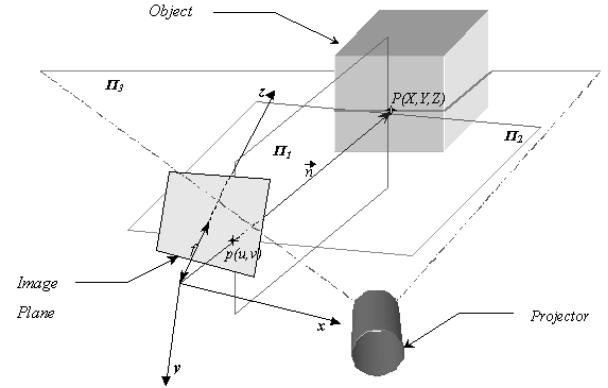


Figure 1 - Generic structured light 3D scanner [1].

It becomes obvious that the information extracted from the two-dimensional image-plane of a projective camera is not sufficient to completely reconstruct a three-dimensional scene. Thus, at least one additional restriction is needed to establish an univocal correspondence between the 3D point and its projection on the image-plane. There are several ways of achieving this goal, one of which being the use of *projected structured light*.

It is clear that the light source is in fact projecting a plane of light, which can be mathematically represented by equation 1.

$$\begin{bmatrix} a & b & c & d \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = 0 \Leftrightarrow \Pi_3 \mathbf{P}' = 0 \quad (1)$$

Using the geometry involved in the perspective projection of the world onto the camera's image-plane and the light-plane's restriction of equation 1, the 3D point in the scene can be uniquely related to its two-dimensional projection point in the image-plane of the camera by triangulation using:

$$\begin{cases} \Pi_1 \mathbf{P}' = 0 \\ \Pi_2 \mathbf{P}' = 0 \\ \Pi_3 \mathbf{P}' = 0 \end{cases} \quad (2)$$

For a more complete explanation of these principles, please refer to [1].

In figure 2, some examples from our preceding research project with a non-portable structured light 3D scanner are presented. It shows clearly that, to achieve complete 3D

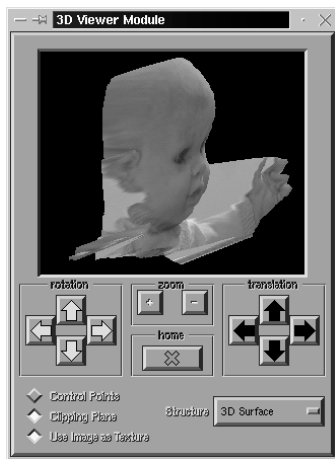


Figure 2 - Example of a 3D reconstruction using a laser rangefinder from one point of view [1] – the necessity for more information to complete it is clear.

scene reconstructions, it will be necessary to integrate more data taken from other points of view.

B. 3D Information Integration

In our present task of building a system with the largest possible degree of autonomy, the greatest challenge was to address data integration from several acquired sets of 3D measurements. Due to several performance issues [2], we have decided to implement a system using a hybrid attitude and position estimation procedure, where both information intrinsic to the acquired data and extrinsic information measured from sensors are used to achieve data integration. We will describe the principles behind this implementation in the following subsections.

B.1 Attitude and positioning

There are basically three ways of measuring the system's position and attitude:

- the use of contrasting landmarks, strategically placed so as to execute triangulations that allow determination of position and attitude changes;
- the transmitter/receiver approach, where a transmitter is placed somewhere on the scene to be recovered while a receiver, placed on the system, is registering relative position and orientation changes using its readings from the transmission;
- the self-contained approach, where the system obtains its position and attitude measurements through a configuration that is independent from the scene to recover.

Any transmitter/receiver system using the time-of-flight or phase-based properties of any physical reality, such as electromagnetic or sound waves, can be used to implement the second approach, as can magnetic systems using physical properties such as the Hall effect.

However, to implement self-contained position and attitude measuring systems, it is necessary to resort to more specific sensors.

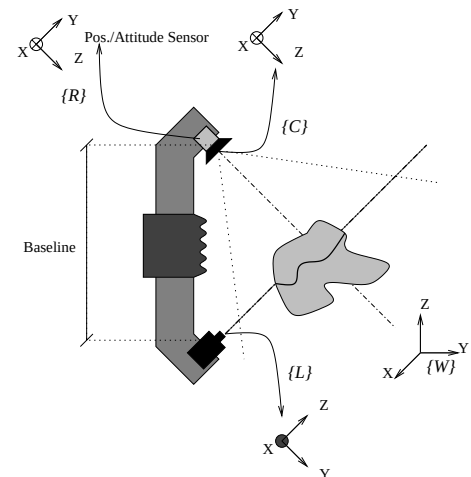


Figure 3 - Handheld architecture schematics.

Gyroscopes and accelerometres are known as *inertial sensors* since they exploit the property of inertia, i.e. *resistance to a change in momentum*; to sense angular motion in the case of the gyro, and changes in linear motion in the case of the accelerometer. Linear velocity and position, and angular position, can be obtained by integration [3].

At the most basic level, the inertial system simply performs a double integration of sensed acceleration to estimate position. Clearly, time integration of measured accelerations corrupted by measurement errors will undoubtedly result in drift error in estimated position that will accumulate over time. This means that, either one must compensate somehow for that drift, compensation proved in practice to be virtually impossible to sustain, or one must rely exclusively on 3D registration to deal with translations; drift effects for rotations, on the other hand, can be effectively compensated [3].

Another sensor which can be included to reset the accumulated drift in the attitude computation is the *3D magnetic compass*, since a good source for absolute orientation of a mobile system is the earth's magnetic field. However the sensors' accuracy is limited in the situation when the earth's magnetic field is distorted [3].

B.2 3D registration

Consider an overview of our handheld system — a camera and a laser projector mounted in fixed position on a rigid structure that will be moved by hand so as to completely scan the 3D scene to recover. At any time, as can be inferred from figure 3, the camera and the laser projector's referentials are fixed relatively to each other and can thus be locally calibrated beforehand.

Image frames of the scene will be grabbed in sequence as the scan is performed and a set of 3D profiles will be determined from each intersection of the projected light-plane with the 3D surfaces on the scene. These profiles, designated as $P(t_n)$ for each sampling time $t = t_n$, represent, in fact, a set of ordered points $x(t_n)$ sampled from those surfaces. Contrary to what happened with the fixed positioned laser scanner of our previous work, the handheld scanner and its local referentials are moved and are rotated during

the scanning process. Since no scaling or reflections are involved, each position and attitude change of the system between consecutive sampling instants can be modelled as a *rigid-body transformation* [4]-[6], a combination of a rotation with a translation expressed by

$$\mathbf{x}(t_n) = \mathbf{R}_{n-1,n}\mathbf{x}(t_{n-1}) + \vec{\mathbf{t}}_{n-1,n} \quad (3)$$

where $\mathbf{x}(t_n)$ corresponds to any sampled point at sampling time $t = t_n$ and $\mathbf{R}$ and $\vec{\mathbf{t}}$ are, respectively, the rotation matrix and the translation vector that compose the transformation, with coordinates referred to any of the system's local referentials (for example, figure 3's $\{L\}$ or $\{C\}$).

Evidently, for complete integration of acquired data into a coherent reconstruction of the scene, computation of all such rigid transformations of any chosen system's local referential is needed so as to perform *3D registration*. The three-dimensional registration process has been defined in several ways depending on the implementation (check [7], [4], [8] for examples), but can be formalised in a more generic form as the *transformation of sets of three-dimensional measurements into a common coordinate system*.

A 3D registration technique can be described mainly by how it addresses 3 crucial problems [4], [8], [6]:

1. Choice of feature space (i.e., type of 3D measurements used: points, curves, planes, voxel images, etc.);
2. Choice of transformation (including rotation representation — by an orthonormal matrix, by quaternions, etc.) and, consequently, of the objective-function to be optimised and its parameter vector;
3. Feature matching (determining correspondences between features is mandatory, since registration will be impossible without them) and global optimisation.

A preferred registration method seems to be the *Iterative Closest Point* algorithm (ICP) and its variations, a technique devised by Besl and McKay and discussed in [8]. It involves a series of line searches in a 7-parameter space spanned by a rotation quaternion and translation vector that model the corresponding rigid-body transformation, guaranteeing local convergence which will be a good solution if given an appropriate initial state for global matching [4], [8]. If the feature set is chosen to be a set of 3D points, for example, the original ICP algorithm operates as follows, given the point sets P and Q where $P \subset Q$ [7], [8]:

Nearest point search: For each point $\mathbf{p}$ in P , determine the closest point $\mathbf{q}$ on Q (supposedly a match)

Compute registration: Evaluate the rigid-body transformation that solves the least-square distance minimisation problem stated in equation 4 [6].

$$\min \sum_{i=1}^N \left\| \mathbf{R}_{i-1,i}\mathbf{x}_{i-1} + \vec{\mathbf{t}}_{i-1,i} - \mathbf{x}_i \right\|^2 \quad (4)$$

Transform: Apply transformation to all points in set P

Iterate: Repeat previous steps until convergence

III. SYSTEM ARCHITECTURE

On figure 3, we present the handheld scanner's architecture. A rigid structure with a straight angle boomerang shape has the camera mounted on one side and the laser projector on the other; a handling grip is placed at the centre of mass of the structure. Coupled with the laser projector is the attitude and positioning device, using inertial and magnetic sensors. Also on figure 3, all referentials directly or indirectly involved on the scanning process are represented: the camera's referential, $\{C\}$, the laser projector's referential, $\{L\}$, the position and orientation measuring sensor's referential, $\{R\}$, and the world's absolute referential $\{W\}$.

IV. PRESENT STATE OF IMPLEMENTATION

We will now describe several implementation issues we are currently addressing, namely system calibration and 3D information integration issues.

A. Calibration issues

For our handheld 3D scanner project, the first issue we addressed was the optimisation of our calibration procedures. Complete system calibration involves determining the position and orientation of the camera, the laser projector and the position and orientation measuring sensor relatively to the chosen world coordinate system [9]. In other words, the calibration procedure encompasses the determination of all coordinate systems that are directly involved in the handheld 3D scanner's operation and, consequently, all rigid transformations between them (please refer to figure 3).

With this purpose in mind, a robust camera calibration software package was developed using Intel's Open Source Computer Vision Library [10]. The OpenCV's calibration method, mainly based on [11], is an iterative algorithm applied to the pinhole model with radial and tangential distortion coefficients. It uses a chessboard pattern to supply 3D points with well-known coordinates and operates as follows [10]:

1. Find homography for all points on a series of images.
2. Initialise intrinsic parameters; set distortion parameters to 0.
3. Find extrinsic parameters for each image of calibrating pattern.
4. Make main optimisation by minimising error of projection points with all parameters.

As this procedure allows determining the rigid transformation between the world's referential and the camera's referential, ${}^W\mathbf{T}_C$, for each point of view involved in camera calibration, it is possible to use that particular information to determine the sensor's position and orientation through the rigid transformation ${}^R\mathbf{T}_C$, which is in turn determined using

$${}^W\mathbf{T}_C = {}^W\mathbf{T}_T \cdot {}^T\mathbf{T}_R \cdot {}^R\mathbf{T}_C \quad (5)$$

where $\{T\}$ is the transmitters referential, if using a transmitter-receiver approach, or the startup referential, if using a self-contained approach. ${}^T\mathbf{T}_R$ is measured directly by the sensor, making thus ${}^W\mathbf{T}_T$ and ${}^R\mathbf{T}_C$ the unknown

matrices to determine (totalling 24 unknowns, 12 of which linearly independent). It is therefore necessary for calibration to obtain at least 2 different equations as stated in equation 5 to determine these unknowns, which is possible by measuring and computing the remaining variables through different positionings of the system relatively to a calibration plane containing the aforementioned chessboard pattern.

Through the knowledge of ${}^W\mathbf{T}_C$ it is also possible to obtain the rigid transformation between the laser's referential and the camera's referential, ${}^C\mathbf{T}_L$, using a procedure which is very similar to what is stated in [9] for the single-wall phantom, with

$$\begin{bmatrix} X_L \\ 0 \\ Z_L \\ 1 \end{bmatrix} = {}^C\mathbf{T}_L^{-1} \cdot {}^W\mathbf{T}_C^{-1} \cdot \begin{bmatrix} X_C \\ Y_C \\ 0 \\ 1 \end{bmatrix} \quad (6)$$

where ${}^C\mathbf{T}_L$ is the unknown matrix, and X_C , Y_C and ${}^W\mathbf{T}_C$ are the known variables/matrix. Bear in mind that these matrices are known to be invertible, since they represent rigid transformations: these are always invertible in their homogeneous, linearised form.

The laser plane representation as stated in equation 1 is determined easily, since the laser's referential has been chosen so as to coincide the laser plane with its XZ plane; that being so, the plane's coefficients affected to the camera's referential are readily computed through their rigid transformation with ${}^C\mathbf{T}_L$, knowing that its coefficients affected to the laser's referential, are, in fact, $a = c = d = 0$ and $b = 1$.

B. Other issues

Several studies concerning the system's set-up, the registration methods and positioning devices are being made. Some important issues have already become clear from these studies:

1. For the registration methods to work, *profiles must overlap*, i.e., a same region of the 3D scene must be scanned at least twice; since profiles are finite, these must also cross inside the scene's boundaries (see [2]) — this is due to the matching features constraint, that implies that, given two feature sets, one must be a subset of the other.
2. In order to simplify the model for registration and to adapt it for use with the positioning and attitude measurement devices, the rigid-body transformation expression can be reformulated in a manner which decouples translation computation from that of rotations by referring the coordinates to the respective centroids of each point set (see [4]). The translation from sets A to B of matching points would then be given by:

$${}^B\mathbf{t}_A = \frac{1}{N} \sum_{j=1}^N {}^B\mathbf{x}_j - {}^B\mathbf{R}_A \frac{1}{N} \sum_{j=1}^N {}^A\mathbf{x}_j \quad (7)$$

3. The 3D scenes must be quasi-static for the system to work properly, making scanning of organic subjects, such as human bodies, difficult to achieve.

V. DISCUSSION AND CONCLUSIONS

Practical areas of use for 3D scanning systems, as described in this article's introductory section, will push towards a revolution in this field of research and small, portable, fast, reliable and cost-effective systems will be available as technologies advance. The current challenge is, thus, to optimise solutions to bring out the best of each implementation. This task, however, is anything but trivial. Perhaps each implementation will only work as intended for particular instances of the 3D reconstruction problem, therefore becoming tailor-made solutions for particular types of scanning.

It is thus the authors' conviction that a lot of experimental work in this field is still to come.

REFERENCES

- [1] João Ferreira and Jorge Dias, "A 3D Scanner - Three-Dimensional Reconstruction From Multiple Images", in *Proc. Controlo 2000 Conf. on Automatic Control*. University of Minho, Portugal, 2000, pp. 690–695, Student Forum.
- [2] P. Hébert and M. Rioux, "Toward a hand-held laser range scanner: Integrating observation-based motion compensation", in *Proceedings of IS&T/SPIE's 10th Annual Symposium on Electronic Imaging (Photonics West); Conference 3313: Three-Dimensional Image Capture and Applications*, San Jose, Ca, USA, January 1998.
- [3] Jorge Lobo, Lino Marques, Jorge Dias, Urbano Dias, and Anibal T. de Almeida, *Sensors for Mobile Robot Navigation*, pp. 50–81, Number 236 in Autonomous Robotic Systems - (Lecture Notes in Control and Information Sciences). Springer-Verlag, 1998, ISBN 1-85233-036-8.
- [4] Michel A. Audette, Frank P. Ferrie, and Terry M. Peters, "An Algorithmic Overview of Surface Registration Techniques for Medical Imaging", *Medical Image Analysis*, vol. 4, no. 3, pp. 201–217, September 2000.
- [5] Jacques Feldmar and Nicholas Ayache, "Rigid, Affine and Locally Affine Registration of Free-Form Surfaces", Research Report 2220, Institut National De Recherche En Informatique Et En Automatique (INRIA), March 1994, Project Epidaure.
- [6] O. D. Faugeras and M. Hebert, "The Representation, Recognition and Locating of 3-D Objects", *The International Journal of Robotics Research*, vol. 5, no. 3, pp. 27–52, Fall 1986, Massachusetts Institute of Technology.
- [7] A. Hilton, "Model Building from 3D Surface Measurements", Web page - http://www.ee.surrey.ac.uk/Research/VSSP/3DVision/model_building/model.html, December 1997, Tutorial.
- [8] Paul J. Besl and Neil D. McKay, "A Method for Registration of 3-D Shapes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, February 1992.
- [9] R.W. Prager, R.N. Rohling, A.H. Gee, and L. Berman, "Automatic calibration for 3-D free-hand ultrasound", Tech. Rep. CUED/F-INFENG/TR 303, Cambridge University Department of Engineering, September 1997.
- [10] Intel Corporation, USA, *Open Source Computer Vision Library*, 1999–2001, Reference Manual.
- [11] Zhengyou Zhang, "A Flexible New Technique For Camera Calibration", Tech. Rep., Microsoft Research, 2000.