From accurate range imaging sensor calibration to accurate model-based 3-D object localization

Guillaume Champleboux[1], Stéphane Lavallée[1], Richard Szeliski[2], Lionel Brunie[1]

[1]TIMB - TIM3 - IMAG, Faculté de Médecine de Grenoble 38700 La Tronche, France
tel: (33) 76 63 71 47, fax: (33) 76 51 86 67, e-mail: lavallee@timb.imag.fr

[2]Digital Equipment Corporation, Cambridge Research Lab
One Kendall Square, Bldg. 700, Cambridge, MA 02139. e-mail: szeliski @crl.dec.com

Abstract

This paper examines the registration of multiple 3-D data sets obtained with a laser range finder. We propose first a new sensor calibration technique based on the conjunction of a mathematical camera model: the (NPBS) model N-planes B-Spline *with an accurate mechanical calibration setup. Second, we develop an algorithm for recovering the rigid transformation (rotation and translation) between two sets of 3D coordinates obtained by the calibrated imaging range sensor. Its input are sets of points lying on the surface of the object. This input is converted into an* octree-spline *which allows us to quickly compute point to surface distances. A robust nonlinear least squares minimization technique then finds the optimal pose by minimizing the sum of squared distances between the two sets of 3D coordinates. We have applied this localization algorithm to matching human face surfaces and have obtained highly accurate results.*

1 Introduction

Estimating the pose of a 3D object from range data is a classical problem in computer vision. Given a model of the object in a coordinate system $\text{Ref}_{\text{model}}$ and given some 3D coordinates of surface points in a coordinate system $\text{Ref}_{\text{sensor}}$, the problem is to estimate the 6-component vector $\mathbf{p}$ that defines the rigid body transformation $\mathbf{T}$ between $\text{Ref}_{\text{sensor}}$ and $\text{Ref}_{\text{model}}$ (rotation and translation). If a matching score can be provided at the end of the parameters estimation, such a localization scheme can be integrated into a more general recognition algorithm. However, the model-based 3D localization problem itself has many applications. For instance, in the medical and surgical fields, various imaging sensors can provide specific information for a same patient (e.g., Computed Tomography (CT), Magnetic Resonance Imaging (MRI), PET and 3D Ultrasound images). A real need is to register all of these 3D images in the same reference system, and to then link these images with the operating instruments such as guiding systems or robots [1]. To achieve this goal, one possibility is to use some anatomical surfaces as references in all these images. In some cases, it becomes necessary to use a laser range finder fixed to one imaging device to acquire the skin surface of a patient, and then to registrate this reference surface with the skin surface segmented on another imaging device [2].

The method described in this paper to solve the 3D localization problem is very general. Keeping in mind the medical applications, our main requirement is to achieve the best accuracy possible for the 6 rigid body transformation parameters. Thus, *accuracy* is the predominant theme of this paper. Moreover, our method needs to be very general and applicable to a wide class of objects. For instance, no global constraint can be introduced about the shape of a smooth object such as a human face (e.g., no planar surfaces, no symmetries, no global modelling with superquadrics). Thus, the second requirement is to perform the matching process for arbitrary free-form smooth surfaces. For our medical applications, the other requirements are to compute an estimate of the uncertainties in the 6 parameters, and to perform the matching process in a reasonable time. Obviously, it would be very useful to meet these requirements in many other applications of model-based 3D localization such as pattern recognition, motion estimation, and robot manipulation.

A general model-based sensing process can be divided into three stages: first, the construction of a 3D model for a given object, second, the measurements of properties of the object in the environment, and third, the interpretation of these measurements. To our knowledge, no method that deals with these three steps meets the previously defined requirements with respect to accuracy and generality. Usually, authors focus on only one of these steps. Because accuracy, which is our first requirement, must be controlled at each stage, our method has to deal with these three levels of sensing. In this paper, we present new algorithms for each of these three steps, we present some results about the integration of these algorithms, and we show that our defined requirements are met.

0-8186-2855-3/92 $3.00 ©1992 IEEE

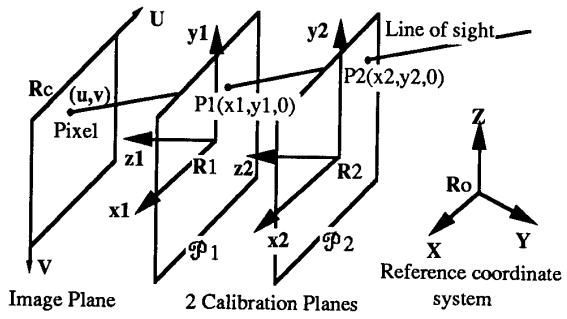

Figure 1: N-planes method for camera calibration

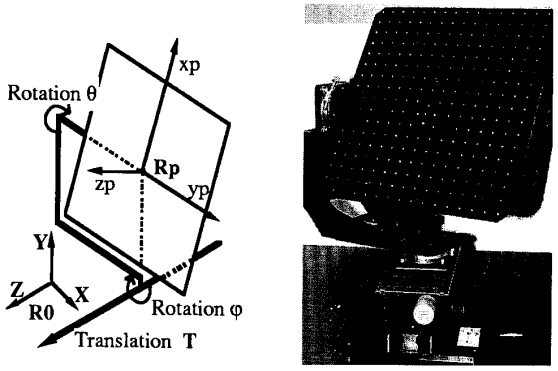

Figure 2: calibration setup made of an articulated plate with 3 d.o.f.

2 Range Imaging Sensor Calibration

Among many sensors that provide 3D coordinates, we have chosen to focus on laser range finders [3]. Our sensor is made of a video camera which observes the projection of a laser plane onto an object. It seems surprising that very few papers describe an accurate calibration method for these systems, and classically many authors use approximations. [4] presents a general calibration method, but its drawback is to propagate the camera calibration errors to the determination of the laser plane. Thus we have purposed a new accurate calibration method for both the camera and the laser plane.

2.1 Previous work on camera calibration

A complete camera calibration review has been proposed by [5]. Two different main categories can be distinguished. (A) First, some methods consist in determining a perspective matrix using homogeneous coordinates [6, 7]. Some additional parameters are then added to this simple pin-hole model in order to take various complex physical phenomena into account [7]. (B) Another approach, much less known uses only a mathematical model for estimating the mapping between the 2D pixels and the 3D projection lines. The method called the two-planes method was first proposed in [8] and then extended in [9]. The principle of the basic method is described on figure 1: each pixel $(u\ v)$ of the image plane is linked to 2 points $\mathbf{P}_1$ and $\mathbf{P}_2$ that belong respectively to two real calibration planes: $\mathcal{P}_1$ and $\mathcal{P}_2$. Each plane $\mathcal{P}_i$ is associated with a coordinate system R_i such that the plane equation is $z = 0$ in the coordinate system R_i. From external measurements, these plane equations and coordinate systems R_i are known in a reference coordinate system Ref_0. Thus, the calibration process consists in estimating the transformations between image coordinates $(u\ v)$ and coordinates of points $\mathbf{P}_1 = (x_1\ y_1\ 0)^t$ in R_1 , and $\mathbf{P}_2 = (x_2\ y_2\ 0)^t$ in R_2. Three solutions have been proposed for these transformations:linear,quadratic, and local. In both categories (A) and (B), the need to estimate *local distorsion* has been pointed out by several authors and the most powerful approach has been to consider linear approximations in triangular or rectangular patches delimited by calibration points [7, 8].

In next section, we propose to extend and generalize these piecewise linear approximations by using regularization theory.

2.2 The NPBS model for camera calibration

We suppose that any pixel $(u\ v)$ of the image plane can be linked to a 3D line of sight, and choose to define a line of sight by two points $\mathbf{P}_1$ and $\mathbf{P}_2$ that belong to 2 planes $\mathcal{P}_1$ and $\mathcal{P}_2$. Thus, the most general formulation of camera calibration is to estimate a function from R^2 (image coordinates) to R^4 (the space of 3D lines has a dimension 4). Thus our method can be seen as an extension of the two-planes model [8]. Using notations presented on figure 1, the problem is to estimate a function $F : R^2 \rightarrow R^4$ defined by

$$(x_1, y_1, x_2, y_2) = F(u, v). \qquad (1)$$

For convenience, we introduce the four functions $F_{x_1}, F_{y_1}, F_{x_2}, F_{y_2}$ defined by

$$x_i = F_{x_i}(u, v) \quad y_i = F_{y_i}(u, v) \quad i = \{1 \ldots 2\} \qquad (2)$$

In the calibration process, given a data set of pixels (u_j, v_j) linked to calibration points (x_{1_j}, y_{1_j}) for $\mathcal{P}_1$, and given a data set of pixels (u_k, v_k) linked to calibration points (x_{2_k}, y_{2_k}) for $\mathcal{P}_2$, the problem is to interpolate the data with functions $F_{x_1}, F_{y_1}, F_{x_2}, F_{y_2}$. Here, we see that camera calibration comes to a classical problem of approximation of some functions $f : R^2 \rightarrow R$, given irregularly distributed data points. For our work, we use the spline theory and bicubic B-Spline functions to represent the four functions of (2), This model introduced in this paper is called the N-Planes B-Spline model and is abbreviated as NPBS. Because we need a physical representation of theses planes, a calibration set-up has been conceived. It is a metallic plate, containing 15x15 holes illuminated by the back side, mounted on an articulated system with three degrees of freedom (dof): two rotations defined by angles θ and ϕ and one translation measured by a variable T. The transformation $[\mathbf{H}(\theta, \phi, \mathbf{T})]$ between coordinates $(X\ Y\ Z)$ expressed in R_p and $(x\ y\ z)$ in Ref_0 (associated to the base of the articulated system) is known

through accurate mechanical assembly. A calibration plane $\mathcal{P}_i$ $i = \{1 \ldots 2\}$ defined by its 3 d.o.f. (θ_i, ϕ_i, T_i), and associated to the transformation $[\mathbf{H_i}]$ gives for any pixel (u, v) a point $\mathbf{P}_i = (x_i\ y_i\ 0)^t_{R_i}$ Adding B-Spline functions introduced in (2) the coordinates of $\mathbf{P}_i$ are given in the absolute coordinate system Ref_0 by :

$$(X_i\ Y_i\ Z_i)^t_{Ref_0} = [\mathbf{H_i}](F_{x_i}(u,v)\ F_{y_i}(u,v)\ 0)^t_{R_i} \quad (3)$$

Finally, each pixel (u, v) is linked to a line of sight D given by 2 point $\mathbf{P}_1$ and $\mathbf{P}_2$ in Ref_0. We can use more than two planes and associate to the pixel (u, v) the line that fits all points $P_1, P_2, \ldots P_N$ (with a least squares approximation) of N planes. This scheme defines the *NPBS model.*

Using the calibration set-up we collect calibration data $(x_{i_k}, y_{i_k}, u_{i_k}, v_{i_k}), k = \{1 \ldots K\}$ for each calibration plane $i = \{1 \ldots 2\}$ (typically, $K = 120$). The estimation of the NPBS model with the calibration data is described in [10]. Using spline theory, this process estimates the 4 functions $f(u, v)$, f beeing $F_{x_1}, F_{y_1}, F_{x_2} or F_{y_2}$, with :

$$f(u,v) = \sum_{m=-3}^{n-1} \sum_{l=-3}^{n-1} \alpha_{m,l}\ B_m(u)\ B_l(v) \quad (4)$$

where n is the number of subdivisions on the u and v length, B_m and B_l are the cubic spline components of the Base Spline [11]and α the matrix of $(n+3)^2$ coefficients α_{ml} to be computed. Setting our problem as a classical regularization problem, the B-Spline function f minimizes a functional J defined by :

$$J(f) = S(f) + \rho \sum_{k=1}^{K} d_k(f, (u_k, v_k)) \quad (5)$$

where the first term of $J(f)$ tends to smooth the surface, while the second term tends to interpolate the data.

2.3 Range imaging sensor model and calibration setup

Succesive motorized regular displacements measured by d of the laser plane on the object viewed by the camera give a set of points (u, v, d) which are transformed into 3D coordinates (X, Y, Z) by triangulation (figure 2.3a). The calibration problem is to find the transform $G : \mathrm{R}^3 \rightarrow \mathrm{R}^3$ defined by: $(X, Y, Z) = G(u, v, d)$. All this can be resumed by calculating the laser-plane line of sight intersection. To obtain the equation of the laser plane in the coordinate system of the articulated calibration device Ref_0, we tune manually (θ, ϕ, T) until the laser plane and the calibration plate meet exactly. The intensity of the light must be maximal at the bottom of the device (figure 2.3b).

2.4 Results for camera and laser range finder calibration

2.4.1 Results for camera calibration

To obtain results about accuracy for each model, we first calibrate the camera (Pers, Line, Quad, Loca are

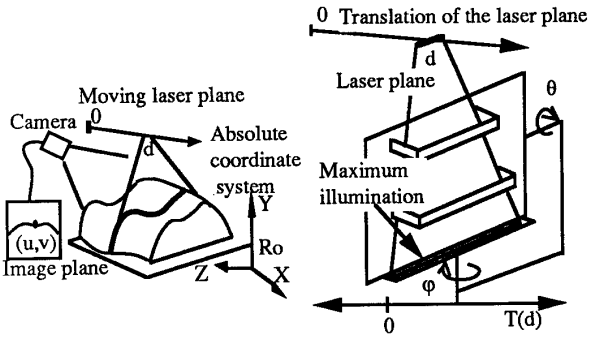

Figure 3: Active triangulation: (a) configuration of a range image sensor with a moving laser plane . (b) Calibration of the laser plane tuning (θ, ϕ, T).

Method:	Pers	Line	Quad	Loca	NPBS
$\Sigma(inmm)$	2.19	3.25	1.63	0.67	0.46
$\sigma(inmm)$	0.63	0.90	0.39	0.20	0.15

Table 1: Test of 5 camera models. Eight experiments have been conducted for each camera model. For each experiment, 5 data planes of 120 data points have been acquired (3 calibration planes, 2 test planes).

respective abbreviation for the perspective transformation model, the 2 Planes linear, quadratic, local, model) Then, we compute the mean and max distances σ and Σ between points theoricaly (using arbitrary positions of the metallic plane with the calibration device) and pratically (with the camera model) determined. Results are reported in table 1, they show that the NPBS method improves the accuracy noticeably.

2.4.2 Results for laser range finder calibration

For eight random positions of the metallic plate we record the plate configuration (θ, ϕ, T) and compute the true equation of this plane $\mathcal{P}$. Using the range imaging sensor calibrated with the NPBS method, we collect a set of point $(X_i, Y_i, Z_i), i = \{1 \ldots M\}$ (M is several thousands, the data are collected in a cube of $30cm^3$) which belong to the surface of the metallic plate. For each of these planes we compute the maximal and mean distance from the M points (X_i, Y_i, Z_i) to the plane $\mathcal{P}$ which are $\Sigma_r = 1.01mm, \sigma_r = 0.54mm$, We calculate the maximal an mean angle between the two planes $\mathcal{P}_{fit}$ and $\mathcal{P}$ and find respectively 0.30° and 0.18°.

2.5 Conclusion about sensor calibration

The defined NPBS model can be applied to any camera, to digitized radiography and to range imaging sensors. It is an alternative to the complex modelling of cameras and frame grabbers with parameters that have physical meanings. The NPBS model does

not require any a priori knowledge about the sensor: the only assumption is that each pixel is associated to a line of sight. We have combined this mathematical modelling with an accurate calibration setup based on an articulated plate with 3 degrees of freedom. This device enables us to acquire accurate calibration data sets. Using a specific mechanical feature, the camera and a laser plane are calibrated independantly, without propagating uncertainties. The NPBS calibration method has been tested for both cameras and range imaging sensors. The achieved accuracy meets the requirements for the intended applications (relative accuracy for camera 1/2000 and relative accuracy for range imaging sensor 1/500). Most importantly, the NPBS method results in an accuracy which is similar to the range imaging sensor resolution.

3 Modelling of surface points using an Octree-Spline

After accurate acquisition of range data, the second step of model-based sensing is to define a model. This section describes previous work in this field and our choice.

3.1 Previous work

In model-based recognition and localization, the choice of a model for the object depends on the matching algorithm that will be used. For this reason, several models have been proposed in order to speed or to simplify the task of matching the data with the model. See [12] for a review. Models that have been used include polyhedrals, generalized cylinders, superquadrics, Extended Gaussian Images (EGI), global invariants Hough transforms, surface patches, and local invariants. For most of these approaches, objects are modeled by graphs of primitives (e.g., edges or vertices, planar patches, polynomial patches). Thus they raise the problem of segmentation into primitives and the problem of correspondence between primitives for the matching process. EGI methods are limited to convex objects. Hough transform methods and methods that match invariants (such as moments) are global and thus they are sensitive to segmentation errors. Such errors occur when only one range image is used or when objects are partially occluded.

3.2 Octree spline

In this paper, a low-level surface representation is introduced: the Octree-Spline. The choice of a low-level representation is motivated by our main objective which is accurate localization, instead of recognition. Here the Octree-Spline enables to build a 3D distance map. The purpose of a distance map is to pre-compute distances $\tilde{d}$ from points to surface and to store them. Such a distance $\tilde{d}(\mathbf{q}, S)$ between $\mathbf{q}$ and S is defined by the minimum of the distances between $\mathbf{q}$ and all points $\mathbf{s}_i$: $\tilde{d}(\mathbf{q}, S) = \min_{\mathbf{s}_i} d(\mathbf{q}, \mathbf{s}_i) = d(\mathbf{q}, \mathbf{s}_{i_{\min}})$, where d is the Euclidean distance. A regular distance map would have about 512^3 or even 1024^3 nodes in order to provide us with the required accuracy. This would raise memory problems, and explains our motivation to introduce an adaptive hierarchical representation : the Octree-Spline.

First, a dense set of n^2 points $\mathbf{s}_i$ lying on the surface of the object is the required initial object representation (almost all surface representations can generate this point representation). The octree-spline is built by an octree decomposition of these surface points, augmented with a continuous 3-D function that approximates the Euclidean distance to the surface. See [13] for a complete description of an octree-spline adapted to solve a 3D/2D matching problem. After the classical octree decomposition of the surface points, the octree is further subdivided (or *restricted*) to ensure that two nodes which are neighbors differ in size by at most a factor of 2, then distances $\tilde{d}(\mathbf{q}, S)$ are computed at each corner of terminal nodes in an exhaustive manner, and finally distance discontinuities between neighboring nodes are eliminated. The intuitive idea behind this geometrical representation is to have more detailed information (i.e., more accuracy) near the surface than far away from it. This representation combines the advantages of adaptive spline functions and hierarchical data structures.

After the previous steps have been performed, $\tilde{d}(\mathbf{q}, S)$ can be computed for any point $\mathbf{q}$ using a trilinear interpolation of the 8 corner values $\tilde{d}_{ijk}$ of the terminal node N that contains the point $\mathbf{q}$ (a binary search starting from the root is used to find N). If $(u, v, w) \in [0, 1]^3$ are the normalized coordinates of $\mathbf{q}$ in the cube N,

$$\tilde{d}(\mathbf{q}, S) = \sum_{i=0}^{1}\sum_{j=0}^{1}\sum_{k=0}^{1} b_i(u)b_j(v)b_k(w)\tilde{d}_{ijk} \qquad (6)$$

with $b_l(t) = \delta_l t + (1 - \delta_l)(1 - t)$.

3.3 Construction of an octree spline from a set of range data

Classically, object models are directly inferred from CAD models. But for some applications, the model can be only derived from sensor data. We have considered this latter case. Using the laser range finder described in section 2, we obtain one range image which is made of a set of surface points $\mathbf{s}_i, i = 1 \ldots n^2$. Typically, by moving the laser plane each 2 mm, using a 512×512 camera image, we obtain $n^2 = 5,000$ or $10,000$ points. Then this set of points is used as input of the octree-spline construction.

4 3-D Localization

As previously discussed in section 3.1, many algorithms have been proposed for model-based recognition and/or positioning. Each of these algorithms has its own advantage which is correlated to the choice of a model description. As we have chosen a low-level approach for our model description (the octree-spline), we go on this low-level approach for the localization algorithm. As before, this choice is justified by the requirement to perform the localization with high accuracy for a free-form surface.

4.1 Problem formulation

In this section, we present our formulation of the localization problem as the minimization of an energy. We look for the rigid transformation $\mathbf{T}(\mathbf{p})$

between a surface S known in $\text{Ref}_{\text{model}}$ and represented by an octree-spline and a set of M_P points $\mathbf{q}_i$ known in $\text{Ref}_{\text{sensor}}$. $\mathbf{T}(\mathbf{p})$ is a 4×4 transformation matrix that depends on a 6-components vector $\mathbf{p} = (T_x\ T_y\ T_z\ \phi\ \theta\ \psi)^t$ (3 translations and 3 Euler angles). First, we make the assumption that most of the points $\mathbf{q}_i$ match the surface. In the ideal case, when the correct matching is reached, each point $\mathbf{q}_i$ is transformed in $\text{Ref}_{\text{model}}$ into a point $\mathbf{r}_i$ that is on the surface S, and every distance $\tilde{d}(\mathbf{r}_i, S)$ is zero. This leads us to formulate the matching problem as a least squares minimization of the energy or error function $E(\mathbf{p})$:

$$E(\mathbf{p}) = \sum_{i=1}^{M_P} \frac{1}{\sigma_i^2}[e_i(\mathbf{p})]^2 = \sum_{i=1}^{M_P} \frac{1}{\sigma_i^2}[\tilde{d}(\mathbf{T}(\mathbf{p})\,\mathbf{q}_i, S)]^2. \quad (7)$$

where σ_i^2 is the variance of the noise of the measurement $e_i(\mathbf{p})$. $E(\mathbf{p})$ is the weighted sum of squares of the distances between the points $\mathbf{r}_i$ and the surface S. Given an initial estimate $\mathbf{p} = \mathbf{p}_0$ of the 3-D/3-D transformation parameters, a nonlinear least squares iterative minimization of the energy (or error function) $E(\mathbf{p})$ is performed.

4.2 Least squares minimization

To perform the nonlinear least squares minimization, we use the Levenberg-Marquardt algorithm because of its good convergence properties. In order to compute the gradient and Hessian of $E(\mathbf{p})$, the Levenberg-Marquardt algorithm requires the first derivatives of each $e_i(\mathbf{p})$. For any component p_j of $\mathbf{p}$, we obtain

$$\frac{\partial e_i(\mathbf{p})}{\partial p_j} = [\nabla \tilde{d}(\mathbf{T}(\mathbf{p})\,\mathbf{q}_i, S)] \cdot \left[\left(\frac{\partial \mathbf{T}(\mathbf{p})}{\partial p_j}\right)(\mathbf{q}_i)\right]. \quad (8)$$

Thus, computing the 6-component gradient of $E(\mathbf{p})$ only requires computing the gradient of the octree spline distance $\tilde{d}$ (by differentiating (6)) and computing the 3 derivatives of $\mathbf{T}(\mathbf{p})$ with respect to the 6 parameters $\mathbf{p}$.

The end of the iterative minimization process is reached either when $E(\mathbf{p})$ is below a fixed threshold, when the difference between parameters $|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}|$ at two successive iterations is below a fixed threshold, or when a maximum number of iterations is reached. At this point, we compute a robust estimate of the parameter $\mathbf{p}$ by throwing out the measurements where $e_i^2(\mathbf{p}) \gg \sigma_i^2$ and performing some more iterations [14]. This process removes the influence of the *outliers* which are likely to occur when the model does not describe the whole surface of the object. The threshold for outlier rejection must be fixed according to application-specific knowledge or by experimentation.

Finally, at the end of the iterative minimization procedure, we estimate the uncertainty in the parameters. The covariance matrix $\text{Cov}(\mathbf{p})$ is computed by inverting the final Hessian matrix and eigenvalue analysis of $\text{Cov}(\mathbf{p})$ is performed [15]. Results and experiments on synthetic data for the localization process are shown in[13]

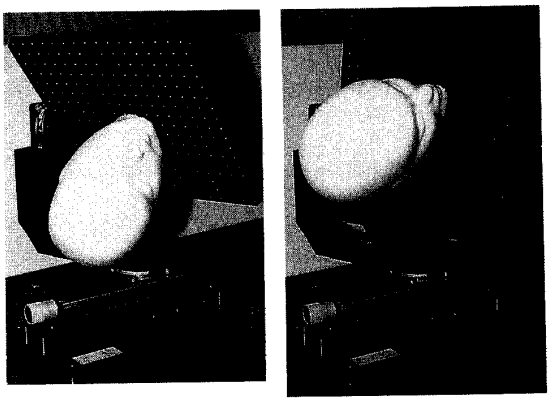

Figure 4: Setup that enables to applicate a known transformation between two range images: the head is stuck on the articulated plate, and each position of the plate is known in the coordinate system Ref_0, thus the transformation between the two positions of the head can be computed in Ref_0.

5 Results for the whole sensing process

5.1 Experiments

The 3 steps of model-based sensing described in this paper have been integrated in order to test the overall accuracy of the localization method. First, a laser range finder has been calibrated as described in section 2. Then a polystyrene head (phantom) has been stuck on the articulated plate described in section 2. For a recorded position (Θ_1, Φ_1, T_1) of the head on the plate, a range image of the face has been acquired. Then a known transformation has been applied to the head. For this second head position recorded by the parameters (Θ_2, Φ_2, T_2), a second range image has been taken. For both images, range data have been given in the coordinate system Ref_0 linked to the base of the articulated device. Thus, the true rigid body transformation $\mathbf{T}^*$ between the head surface in first location and the head surface in second location has been calculated in Ref_0, by computing the composition of the inverse of the transformation $H(\Theta_1, \Phi_1, T_1)$ with the transformation $H(\Theta_2, \Phi_2, T_2)$. Figure 4 shows the setup that enables this computation of the known transform between two range images.

We then applied our new localization algorithm to the two range images. The first image is used to build an octree-spline that constitutes the model. Then, starting from an initial transformation quite far from the solution, we start our iterative localization algorithm between the range data of the second image and the model. At each iteration k, we compute the error transformation $\Delta\mathbf{T}^{(k)}$ and we extract the norm of

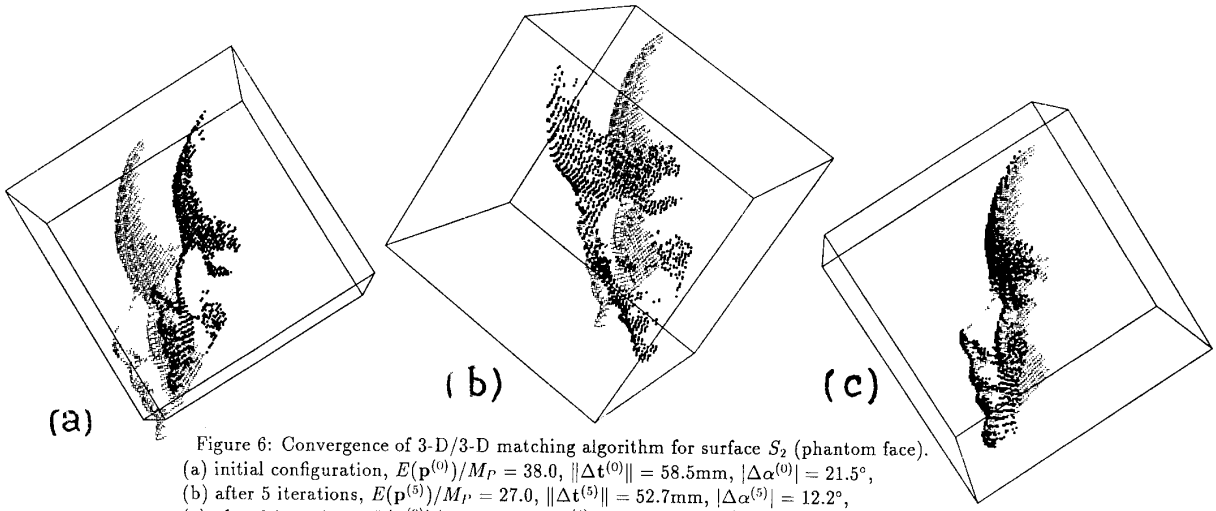

Figure 6: Convergence of 3-D/3-D matching algorithm for surface S_2 (phantom face).
(a) initial configuration, $E(\mathbf{p}^{(0)})/M_P = 38.0$, $\|\Delta\mathbf{t}^{(0)}\| = 58.5$mm, $|\Delta\alpha^{(0)}| = 21.5°$,
(b) after 5 iterations, $E(\mathbf{p}^{(5)})/M_P = 27.0$, $\|\Delta\mathbf{t}^{(5)}\| = 52.7$mm, $|\Delta\alpha^{(5)}| = 12.2°$,
(c) after 9 iterations. $E(\mathbf{p}^{(9)})/M_P = 1.6$, $\|\Delta\mathbf{t}^{(9)}\| = 2.8$mm, $|\Delta\alpha^{(9)}| = 0.7°$.

translation error $\Delta\mathbf{t}^{(k)}$ and $\Delta\alpha^{(k)}$, the angle of the rotation component of $\Delta\mathbf{T}^{(k)}$. The values of $\|\Delta\mathbf{t}^{(k)}\|$ and then we compute $|\Delta\alpha^{(k)}|$ are displayed to monitor the convergence of the algorithm towards the optimal solution. Finally, robust estimation is performed by removing the outliers and by starting some more iterations.

5.2 Final results

This section describes a typical convergence. Figure 5 shows the convergence of the algorithm towards the real solution. The octree-spline is built from 7,500 points (first range image). The second range image gives 2,700 points (all data are not used and a subsample is randomly selected). In about 10s on a DECstation 5000/200, the algorithm reaches a first minimum. At this point, the rotation error is 0.7° and the translation error is $2.9mm$. After the outliers are automatically removed (34 data are suppressed), the algorithm starts again. At the end of the complete algorithm (which takes about 35s), the accuracy is about 0.4° for rotation and $1.2mm$ in translation. Such an accuracy is similar in magnitude to the resolution of the range imaging sensor. Thus the main result is that neither the calibration process with the NPBS method, nor the object modelling with the octree-spline, nor the robust minimization algorithm introduce appreciable inaccuracies. Two other identical experiments have been made, results give the same order of accuracy for both rotation (about 0.5°) and translation (about $1mm$).

5.3 Discussion of the localization algorithm

The localization algorithm has been tested in several experiments and for almost all cases the method has met our requirements. In this section, we discuss when the algorithm fails.

First, if the model is constructed from a set of points that describe one part of the object and if the range data describe another part of the object, obviously the distance minimization gives a wrong estimate. However, if the range data describes an included set of the model surface, the algorithm works perfectly. Moreover, if the number of data points that are not on the surface described by the model is low, these points can be considered as outliers and they are automatically removed by the robust estimation process.

A second failure case occurs when the initial estimate parameters are too far away from the solution. Because the energy function that we minimize is not convex in the whole parameter space, some convergence towards local minima can occur. For our intended medical applications, however, this problem does not appear since we always have a good initial estimate given by a priori knowledge (e.g. the patient's face is always more or less facing up). For other applications, an exhaustive search from randomly spread starting points should give the solution since we expect the energy function to exhibit only a few local minima for most natural objects. This point constitutes one of our future research work.

Our algorithm has been developed to accurately locate one object in a simple scene. Extensions to recognition and extensions to scenes with several objects have not been studied.

Finally, our algorithm can be considered as an extension of the work described in [16], but for more general point to surface distances, for application to range data, and for faster computation times. Similar work has also been recently published in [12], authors propose a very general solution that would probably meet our requirements. However, in our method, only a surface point representation is necessary which makes it very easy to use, and our simple minimization algorithm (based on a continuous distance map) seems much faster (although not the same data have been used).

6 Conclusion

There are many applications of model-based localization in the medical field. The main requirements

common to these various applications are to perform the localization as accurately as possible and to handle smooth free-form objects of arbitrary complexity. Taking into account these requirements, this paper has presented three original methods, one for each level of the sensing process.

First, the importance of accurate data has been stressed and a new method for camera and range imaging sensor calibration has been proposed. The camera calibration problem has been set within a strict mathematical formulation using regularization theory. This lead us to develop the N-Planes B-Splines (NPBS) method. As well, accurate mechanical devices have been designed to obtain accurate calibration data and to calibrate independantly a camera and a laser plane. With this method, experiments show that the accuracy obtained is similar to the sensor resolution. The second contribution of this paper has been to propose a new low-level modelling of surfaces, which is very general. A set of 3D points is used to construct an octree-spline, a new representation that combines an octree with a continuous 3-D distance map. Our third contribution has been to propose a new algorithm for model-based localization using robust nonlinear least squares minimization. Fast and accurate convergence is obtained with this algorithm.

All of these steps have been integrated and accuracy tests have been conducted for the whole sensing process. One set of data is used to constitute the model (octree-spline), and another set of range data is matched with the previous one. The reached accuracy for the full localization process is similar in magnitude to the resolution of the range imaging sensor (about 1 mm in translation and about 0.5 degrees in rotation).

Thus all our requirements for medical applications have been met: the localization is performed for a smooth free-form object, the algorithm does not introduce noticeable inaccuracy, estimation of uncertainties is provided and the convergence time is reasonable.

Acknowledgements

This research is financially supported by Digital Equipment Corporation.

References

[1] S. Lavallee and P. Cinquin. IGOR: Image guided operating robot. In *Fifth International Conference on Advanced Robotic*, Pisa, Italia, June 1991.

[2] S. Lavallee, L. Brunie, B. Mazier, and P. Cinquin. Matching of medical images for computer and robot assisted surgery. In *IEEE EMBS Conference*, Orlando, Florida, November 1991.

[3] P.J. Besl. *Active Optical Range Imaging Sensors.* Springer Verlag, New-York, 1988.

[4] R.C. Bolles, J. Kremers, and R. Cain. Projector-camera range sensing of three-dimensional data. *Machine Inteligence Research Applied to Industrial Automation SRI International*, 12:29–43, 1983.

[5] R. Y. Tsai. Synopsis of recent progress on camera calibration for 3D machine vision. In *The Robotics Review*, pages 147–160. MIT Press, 1989.

[6] T.M. Strat. Recovering the camera parameters from a transformation matrix. *Computer Vision, Graphics, and Image Processing*, pages 93–100, 1987.

[7] O.D. Faugeras and G. Toscani. Camera calibration for 3D computer vision. In *International Workshop on Industrial Applications of Machine Vision and Machine Intelligence*, pages 240–247, Seiken Tokyo Japan, February 1987.

[8] H. A. Martins, J. R. Birk, and R. B. Kelley. Camera models based on data from two calibration planes. *Computer Graphics and Image Processing*, 17:173–179, 1981.

[9] K. D. Gremban, C. E. Thorpe, and T. Kanade. Geometric camera calibration using systems of linear equations. In *IEEE International Conference on Robotics and Automation*, pages 562–567, Philadelphia, Pennsylvania, April 1988.

[10] G. Champleboux, S. Lavallée, P. Sautot, and P. Cinquin. Accurate calibration of cameras and range imaging sensors : the npbs method. In *IEEE International Conference on Robotics and Automation*, Nice France, May 1992.

[11] W. Bohm, G. Farin, and J. Kahman. A survey of curve and surface in CAGD. *Computer Aided and Geometric Design*, 1:1–60, 1984.

[12] P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[13] S. Lavallee, R. Szeliski, and L. Brunie. Matching 3-D smooth surfaces with their 2-D projections using 3-D distance maps. In *SPIE Vol. 1570 Geometric Methods in Computer Vision*, pages 322–336, San Diego, CA, July 1991.

[14] P. J. Huber. *Robust Statistics.* John Wiley & Sons, New York, New York, 1981.

[15] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes: The Art of Scientific Computing.* Cambridge University Press, Cambridge, England, 1986.

[16] C.A. Pelizzari, G.T.Y. Chen, D.R. Spelbring, R.R. Weichselbaum, and C-T. Chen. Accurate 3-D registration of CT, PET, and-or MR images of the brain. *J. Computer Assisted Tomography*, 13(1):20–26, 1989.