Validation of 3-D Registration Methods based on Points and Frames

X. Pennec J.P. Thirion

Epidaure Project
INRIA Sophia-Antipolis
06902 Sophia-Antipolis Cedex, FRANCE
Email: xavier.pennec@sophia.inria.fr
http://www.inria.fr/epidaure/Epidaure-eng.html

Abstract

In this paper, we propose a new method to estimate a rigid transform from a set of 3-D matched points or matched frames, and we concentrate on the analysis of the uncertainty of the estimated transform. The theoretical contributions are an intrinsic model of noise for transformations based on composition rather than addition, a unified formalism for the estimation of both the rigid transform and its covariance matrix for points or frames correspondences, and also a statistical validation method to verify the error estimation, which applies even when no "ground truth" is available. The practical contribution is the validation of our transform estimation method in the case of 3-D medical images, which shows that a precision of the registration, far below the size of a voxel, can be achieved.

1 Introduction

Many algorithms in Computer Vision concern *matching* tasks, whose aim is to find the correspondence between two representations of an object. In the following, we do not discuss matching methods per se, rather the estimation of the transform, or *registration*. Proposed methods generally relies on a least squares estimation, using for example the singular value decomposition (see [AHB87], [Ume91]), or the quaternion representation (see [Hor87]). Some authors, such as [ZH94], concentrate also on the robustness of that estimation.

There are fewer studies, however, dealing with the precision of the estimated transform, which is our main concern here. Early experimental works can be found in [FH84], [Sny89], or [HJLZ89]. A theoretical evaluation of the errors associated to transforms is introduced in [Kan93], mainly for the rotational component. In the present paper, our main contributions are:

- a model of errors in rigid transforms which is not an additive noise, but the composition with an error transform, and a formalism which unifies the handling of the error for the application, composition and evaluation of rigid transforms, in the case of point-to-point or frame-to-frame correspondences (section 2).
- a robust method based on this formalism to estimate both the rigid transform and its associated covariance matrix (section 3)
- an a posteriori method to validate statistically the estimated covariance matrix (section 4).
- the statistical validation of the registration algorithm with both synthetic and real data (3D medical images - section 5).

2 Uncertainty and frames

2.1 Representation of frames

Let $\mathcal{B} = \{O, i, j, k\}$ be the canonical right-handed orthonormal basis of the Euclidean space $\mathbb{R}^3$ and $\mathcal{F} = \{P, i', j', k'\}$ be another right-handed orthonormal basis (i.e. a frame). The rigid transformation from $\mathcal{B}$ to $\mathcal{F}$ can be written with homogeneous coordinates

$$M = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \quad \text{with} \quad R = [i', j', k'] \ , \ t = P - O$$

The matrix M can also be viewed as the matrix representing the coordinate change from $\mathcal{F}$ to $\mathcal{B}$.

Let R be a rotation of angle θ around the unit axis n. Using the rotation vector $r = \theta n$ (see [Aya91, PT95]), we can represent the rigid transformation M by the six dimensional vector $f = (r, t)$.

We note that f represents both the rigid transformation M and the frame $\mathcal{F}$. Other representations could have been chosen for the rotation, e.g. quaternions, Euler's angles, ... but f would still represent the same rigid transformation or feature frame. For convenient notation within the text, we will not distinguish between frames and rigid transformations when there is no ambiguity.

2.2 Operations on frames

Rigid transformations can be composed, inverted and applied to points. The application of a rigid transformation f to a frame f_m is simply the composition $f_s = f \circ f_m$. The compose (operator $\circ$), inverse $(^{(-1)})$ and apply $(\star)$ operations of rotation vectors (see [PT95]) can be easily extended to frames in the following way.

- Application of $f = (r, t)$ to point x:

$$y = f \star x = r \star x + t$$

- Composition of $f_1 = (r_1, t_1)$ by $f_2 = (r_2, t_2)$:

$$f = f_2 \circ f_1 = (r_2 \circ r_1 \ , \ r_2 \star t_1 + t_2)$$

- Inverse of $f = (r, t)$:

$$f^{(-1)} = \left(r^{(-1)} \ , \ r^{(-1)} \star (-t)\right)$$

2.3 Model of noise

Let f be the genuine representation of a frame and $\hat{f}$ its corrupted measurement. We model this error as a composition process: $\hat{f} = f \circ e$ where e is a small rigid transformation around the identity. Due to the asymmetry of the composition, this right error is more suitable than the left error ($\hat{f} = e_l \circ f$) because e_l can be far from identity (see [PT95]).

Note that this noise model is independent of the representation, and therefore intrinsic. It allows the comparison of errors between different frames and provides a uniform way of modeling uncertainty due to the measurement process. Last but not least, the error is itself a transformation which is more tractable. This is significantly different from the usual additive error model $\hat{f} = f + \delta f$, where δf has no meaning since the addition or subtraction of frames is not defined, and moreover completely depends on the representation.

2.4 Uncertainty representation

Let $\hat{f} = f \circ e$ be the *representation* of a probabilistic frame: it is a 6D random vector of probability density ρ. Its expectation is defined by $\bar{f} = E(\hat{f}) = \int \hat{f} \rho(\hat{f}) d\hat{f}$ and its covariance matrix is given by

$$W_f = E\left((\hat{f} - \bar{f})(\hat{f} - \bar{f})^\top\right)$$

We process the probabilistic frame f as the pair $(\bar{f}, W_f)$. A common hypothesis is that the noise is *without bias*, i.e. $E(e) = Id$, which means that $\bar{f} = f$. A probabilistic frame in that case is the pair (f, W_f) and a deterministic frame has a null covariance matrix.

We use for the propagation of covariance matrices the classical first order approximation of Jacobians (see [DW88, ZF92]). Let x be a vector with covariance W_x and h a multi-dimensional function acting on x. The covariance of the vector $z = h(x)$ is, to first order

$$W_z = J_h W_x J_h^\top \quad \text{with} \quad J_h = \frac{\partial h(x)}{\partial x}$$

The propagation of covariance matrices through the operations of application, composition and inversion of frames are derived in [PT95].

2.5 Estimation of a rigid transformation

In this study, we face a registration problem: the estimation of a rigid transformation (f, W_f) from two sets of matched geometric primitives, here points or frames, and a data fusion problem: the merging of several measures (f_i, W_{f_i}). A unified framework for estimation with uncertainty has been developed using the Extended Kalman Filtering (EKF) in [Aya91]. We extend this framework to the estimation from frames correspondences in the following sections.

2.5.1 Estimation from matched points

Assume we have two sets of matched points x_i and y_i, one transformed into the other with a rigid transformation f such that $y_i = f \star x_i$. During the measurement process, those points are corrupted by additive noise: we only measure $\hat{x}_i = x_i + \delta x_i$ and $\hat{y}_i = y_i + \delta y_i$ and what we want to estimate is the transform f and the covariance W_f of this estimation. We assume that the errors δx_i and δy_i have covariance W_{x_i} and W_{y_i}.

Most minimization techniques, and in particular the EKF, rely on the definition of an error vector z_i with covariance W_i that should ideally be *null* for each measure i, and try then to minimize the weighted least squares criterion

$$C = \sum z_i^\top W_i^{(-1)} z_i$$

The error in this case is the classical difference in position $z_i = \hat{y}_i - f \star \hat{x}_i$, and we can use an EKF to estimate (f, W_f).

In the isotropic case (say, when $W_{x_i} = W_{y_i} = \sigma^2 Id$), the criterion reduces to a simple least squares

$$C = \sum \|\hat{y}_i - f \star \hat{x}_i\|^2$$

which can be solved exactly by several techniques (see [Hor87, Aya91]). However, there is no covariance associated with the transformation found.

2.5.2 Estimation from matched frames

We now face the following problem. Assume we have two sets of matched frames $\{f_{m_i}\}$ and $\{f_{s_i}\}$ (m for model and s for scene), transformed one into another with a global rigid transformation f such as $f_{s_i} = f \circ f_{m_i}$. We only have access to their measured values $\hat{f}_{s_i} = f_{s_i} \circ e_{s_i}$ and $\hat{f}_{m_i} = f_{m_i} \circ e_{m_i}$ and we want to estimate the rigid transformation f and a covariance matrix W_f based of that estimate. We assume moreover that we have some knowledge about the measurement process: $e_{m_i} = (Id, W_{m_i})$ and $e_{s_i} = (Id, W_{s_i})$.

Combining the equations and isolating error, we get

$$\hat{f}_{s_i}^{(-1)} \circ f \circ \hat{f}_{m_i} = e_{s_i}^{(-1)} \circ e_{m_i}$$

The left side of this equation is our error vector $e_i = (Id, W_i)$ to be minimized in the criterion

$$C = \sum e_i^\top W_i^{(-1)} e_i$$

We are using here a particular aspect of our representation for frames: identity corresponds to a null translation and a null rotation vector. This allows us to solve this estimation problem using the EKF. The computations needed to linearize the measure equation $e_i = \hat{f}_{s_i}^{(-1)} \circ f \circ \hat{f}_{m_i}$ are given in [PT95]

2.5.3 Fusion of rigid transformations

Consider now a set $\{\hat{f}_i\}$ of measures of the same frame f such as $\hat{f}_i = f \circ e_i$. The goal is to estimate the frame (f, W_f). If we assume an error $e_i = (Id, W_i)$ on each measure, the measure equation is

$$\hat{f}_i^{(-1)} \circ f = e_i^{(-1)}$$

and the corresponding weighted least squares criterion can also be minimized using the EKF.

2.6 Mahalanobis distance and χ^2 test

Considering sets of matched features, we want to sort the matches by saliency and reject outliers in order to improve the robustness of the EKF. Another application of those tools is the clustering of transforms in order to detect and isolate objects (see also [SB]). We derive now the compatibility tests for three different cases.

2.6.1 Two points and a transformation

Let (x, W_x) and (y, W_y) be two measured points and (f, W_f) a rigid transformation. We want to test the validity of the hypothesis $y = f \star x$.

We saw in section 2.4 how to compute $(f \star x, W_{f \star x})$; the covariance on the error vector $z = y - f \star x$ should be zero with covariance $W = W_y + W_{f \star x}$. A 3-D χ^2 test is well suited to verify if z is compatible with its theoretical covariance. The squared Mahalanobis distance μ^2 is

$$\mu^2 = z^\top W^{(-1)} z$$

The statistical test settles that the hypothesis is good if $\mu^2 \leq \varepsilon$, where the threshold ε is set such that, if the hypothesis is true, we will choose it with the probability α. Values of ε for given probabilities α can be found in standard χ^2 tables.

2.6.2 Two frames and a transformation

Let (f_m, W_m) and (f_s, W_s) be two measured frames and (f, W) a rigid transformation. We want to test the hypothesis that $f_s = f \circ f_m$. From section 2.4 we can compute the covariance W on the error vector $e = f_s^{(-1)} \circ f \circ f_m$ defined above and use the squared Mahalanobis distance $\mu^2 = e^\top W^{(-1)} e$ and a 6-D χ^2 test to decide if the hypothesis is true or not.

2.6.3 Two frames

Let (f_1, W_1) and (f_2, W_2) be two measured frames. We want to test the hypothesis that they are measures of the same frame f. Let W be the covariance on the error vector $e = f_1^{(-1)} \circ f_2$. We can use once again the squared Mahalanobis distance $\mu^2 = e^\top W^{(-1)} e$ and a 6-D χ^2.

3 A robust algorithm for transform estimation and accuracy evaluation

The output of a feature-based matching algorithm usually consists in two sets of matched features. We propose a one parameter method (χ^2 test) to estimate the transformation f and its precision W_f, discard outliers, and provide an estimation of the precision of inlier features.

3.1 The bootstrapping process

If the covariance matrix on features is a priori unknown, we start with

- A least squares estimate of f, using only the point positions.
- Estimate the covariance matrix on features.
- Order matches by increasing Mahalanobis distances.

The bootstrap process itself is as follows.

- Estimate the transformation (f, W_f) from the matches using an EKF (in decreasing significance order) until the Mahalanobis distance of the matches becomes larger than a predetermined ε value (this is the χ^2 test to reject outliers).
- Estimate the covariance matrix of all features[1].
- Sort matches using their Mahalanobis distance. Outliers are at the end of the list.

The process is repeated until convergence, or for a preset maximum number of iterations (typically 5).

3.2 Covariance matrix of features

3.2.1 The case of points

Let $\{x_i\}$ and $\{y_i\}$ be two sets of matched points with measured values $\hat{x}_i = x_i + \delta x_i$ and $\hat{y}_i = y_i + \delta y_i$, and an estimate $\hat{f}$ of the transformation $f = (r, t)$ linking them together ($y_i = f \star x_i$).

The error vector $z_i = \hat{y}_i - \hat{f} \star \hat{x}_i$ is, to first order, $z_i = \delta y_i - r \star \delta x_i$ if we consider the error in the estimate of f to be small compared with the error on points (which should be the case if we have enough matched features to estimate it).

Assuming that each measurement error δx_i and δy_i comes from the same (isotropic) process, with covariance W, the covariance on the error vector z_i should be $2W$. We can then estimate W by

$$W_{x_i} = W_{y_i} = W = \frac{1}{2} E(z_i\, z_i^\top) = \frac{1}{2\,N} \sum_i z_i\, z_i^\top$$

It should be noted that neglecting the error on the transformation leads to an overestimation of W, which is an acceptable condition.

3.2.2 The case of frame features

With the same notations as in the previous sections, and neglecting once again the error on the estimate $\hat{f}$ of the transformation f, the error is:

$$e_i = \hat{f}_{s_i}^{(-1)} \circ \hat{f} \circ \hat{f}_{m_i} = e_{s_i}^{(-1)} \circ e_{m_i}$$

Assuming a common underlying process of measurement errors $e_{s_i} = e_{m_i} = (Id, W)$, we can determine that $e_{s_i}^{(-1)} = (Id, W)$, hence $e_i = (Id, 2W)$. It should be noted that the covariances behave here as usual, but only because errors are centered around identity. An estimator of the covariance W of frame features is given by

$$W_{m_i} = W_{s_i} = W = \frac{1}{2} E(e_i\, e_i^\top) = \frac{1}{2\,N} \sum_i e_i\, e_i^\top$$

[1] Computing the covariance matrix with inliers leads to an under-estimation because only small errors are taken into account

3.3 Final Precision evaluation

From $(\hat{f}, W_f)$, we can compute for each point x of the model image (or object) its transform $\hat{y} = \hat{f} \star x$ and the uncertainty $W_y = W_{\hat{f} \star x}$. In order to give a more intuitive information, we can present instead the RMS error expectation at this point: if $y = f \star x$ is the real location of the transformed point x, we shall find it in $\hat{y} = \hat{f} \star x$. The expectation of the squared distance $d^2 = \|y - \hat{y}\|^2$ between them is

$$\sigma_d^2(y) = \mathrm{trace}(W_y)$$

In the case of 3-D medical images, and in order to characterize the precision with a single value, we can estimate the average RMS for the 8 vertices of the image. We call this value the *typical boundary error*.

4 An a posteriori statistical validation

We strongly believe that any estimation method is ultimately statistical. This is why we propose a statistical validation of the estimated covariance matrix associated to the rigid transform, which doesn't requires a *ground truth* transform, but only assumes that the features used have no systematic bias. Note that our technique can be applied directly to registration based on external marker also.

4.1 Description of the method

We consider the registration as a black box, which takes two representations of the same object as input, and gives an estimated rigid transform as output.

Assuming the independence of pairs of matched features, we can subsample the matches in order to get several independent estimates of the searched transformation, and compare them. In particular, we can split the set of matches in two sets and obtain two different estimates of f:

$$f^1 = f \circ e^1 \qquad f^2 = f \circ e^2$$

The two transformations should be very close, and we can study their "distance" $e = (f^2)^{(-1)} \circ f^1 = (e^2)^{(-1)} \circ e^1$ which does not depend of the exact transform f and should be close to identity.

If matches are randomly subsampled to a fixed number M of matches, in order to conserve the independence and a similar distribution of matches in space, then the estimation errors e^1 and e^2 follow the same law of covariance $W(M)$. If moreover both the feature extraction and the registration method have no bias[2] (this means that $E(e^1) = E(e^2) = Id$), then we

[2] The absence of bias for the registration process can be verified on synthetic data, and simple tests can be devised to estimate the feature extraction bias.

can derive that the measured error e has a covariance $W_e(M) = 2W(M)$.

Repeating the experiment with several homogeneous sets of matches, we can estimate the covariance matrix of the estimation of f from M feature matches with

$$W(M) = \frac{1}{2\,n} \sum_{i=1}^{n} e_i\, e_i^\top$$

4.2 Validation of covariance estimates

This method is also used to validate our estimation scheme. We compute the two estimates f^1 and f^2, along with their covariance matrices, and then the error e and its *predicted* uncertainty $W_{e_{pred}}$. The best way to see if the measured values of e are compatible with their predicted uncertainty is to use the Mahalanobis distance of $(e, W_{e_{pred}})$ with identity.

$$\mu^2 = e^\top\, W_{e_{pred}}^{(-1)} e$$

Requirements about the number of features used to compute f^1 and f^2 can then be removed. If the covariance estimation is perfect, then the expected value of μ^2 is 6, the dimension of the transformation space. We use then several sets of matches and estimate the expectation by the average value:

$$\bar{\mu}^2 = \frac{1}{n} \sum \mu_i^2$$

and characterize the adequation of the *predicted* uncertainty of the estimated transform with the *a posteriori* measured values by the *validation index*

$$I_2 = \sqrt{\frac{\bar{\mu}^2}{6}}$$

This index gives indications about how the estimation method under-estimates ($I_2 > 1$) or over-estimates ($I_2 < 1$) the error on the estimated transform. It is a kind of relative error on the error estimation.

In the case of synthetic data with a known transformation f, a direct validation index I_1 can also be computed from the error

$$e = f^{(-1)} \circ f'$$

since by definition the transform estimation is $f' = f \circ e$.

5 Experiments

We present now the validation of our transform estimation algorithm, first for simulated data, and then for real 3D medical image registration.

5.1 Synthetic data experiments

We use a Gaussian error distribution for the positions and the orientations of the feature frames with the same covariances and number of features as for the real experiments (see table 1). This shows a very good fit of the prediction for frames and a slight under-estimation for points. We obtain similar results with a uniform bounded distribution of noise.

	Corner RMS	I_1	I_2
Points	0.120 mm	1.175	1.18
Frames	0.107 mm	0.966	1.025

Table 1: Statistical validation, synthetic case

We also compared the point based and frame based registrations (see figure 1): using frames can then lead to a important improvement in precision, more sensible when the number of matches decreases, and/or with an increasing ratio σ_d/σ_θ.

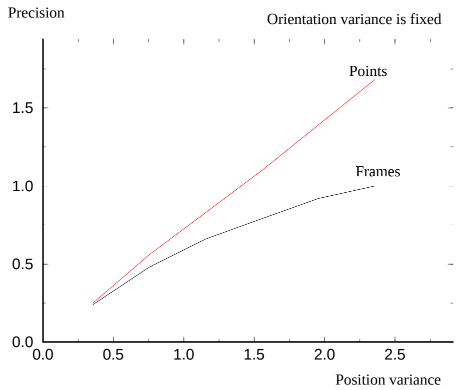

Figure 1: Comparison of the estimation precision using points and frames.

5.2 Real data experiments

We present results from an experiment performed using 3D Magnetic Resonance images (MRI). The images come from the Brigham and Women's Hospital (courtesy of Dr Ron Kikinis), and are part of an extensive study of the evolution of the Multiple Sclerosis disease. The same patient get a complete 3D MR examination several times during one year (24 3D acquisitions with a voxel size of $1\times1\times3mm$): corresponding slices of different 3D MRI are presented in figure 2 and their differences in figure 3.

The matching algorithm is described in [Thi94] and we present here the accuracy obtained by our registration algorithm. Frames are based on the extremal points, defined from differential geometry properties (see [Thi93]), with their associated principal curvature directions and surface normal as trihedron.

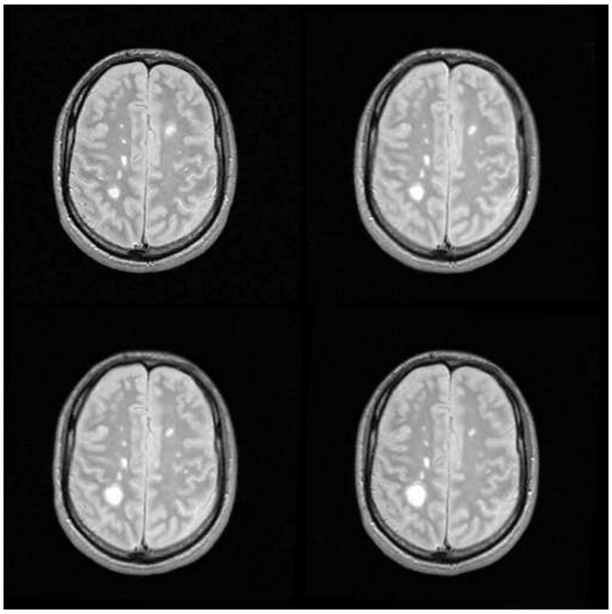

Figure 2: The same slices from four registered 3D MR images of the Multiple Sclerosis (MS) study. Note the evolution of two MS lesions (white spots). One is growing in the anterior left hemisphere, one is shrinking in the right posterior hemisphere. There is two weeks between each 3D acquisition.

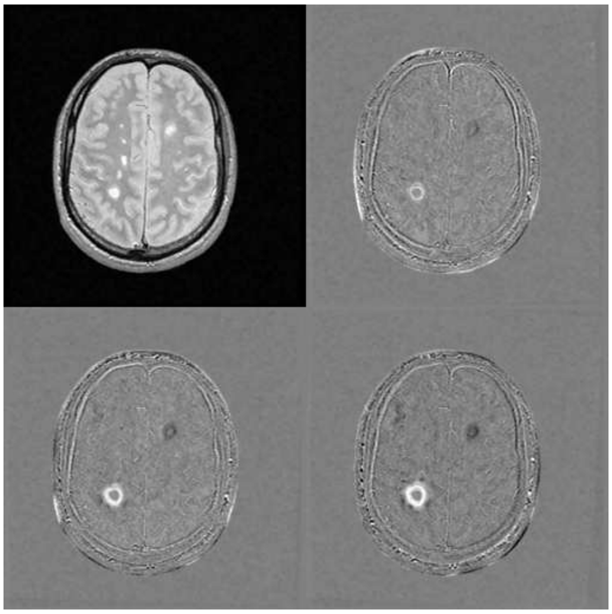

Figure 3: Differences between images after registration, with respect to the first one. The intensity is multiplied by 5, and shifted such that no difference is a gray value. The growing lesion appear as a white disk, and the shrinking lesion as black a disk.

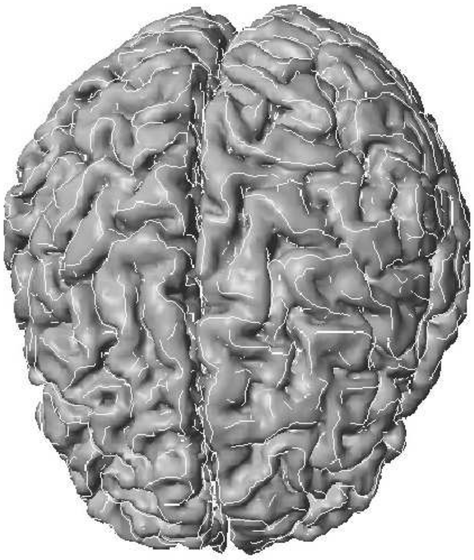

Figure 4: Lines of extremal curvature on the brain circumvolutions. The extremal points, which are used in the registration, are specific points of those lines

5.3 Results

We find a typical boundary RMS error of $\sigma = 0.13\,mm$, which is far below the resolution. The typical predicted covariance on corners is

$$\bar{W}_y = \begin{bmatrix} 0.00496 & -0.00002 & 0.00011 \\ -0.00002 & 0.00473 & -0.00009 \\ 0.00011 & -0.00009 & 0.00707 \end{bmatrix}$$

which reflects the anisotropy of the original images. The validation index is $I_2 = 1.094$, which validates our algorithm for real data.

The registration is enough precise to allow us to visualize, for the first time, the dynamic effect of the lesions on the surrounding brain tissues. With that level of accuracy, improvements can only be obtained through deformable registration.

6 Conclusion

We have developed a new formalism to handle uncertainty for rigid transform evaluation, which consider simultaneously rotation and translation, and which consider residual errors as transforms, and not as an additive noise. This leads to new formulae for the equations of measures which are used within the Extended Kalman Filter framework to evaluate both the transform and its covariance matrix, in the case of 3-D points or 3-D frames matches. This unified formalism is used to elaborate a robust estimation algorithm that reject outliers. In the future, we intend to use it for matching and clustering. We also provide a precise way to validate the predictions of our registration scheme. A practical result is to show that, in the case of 3D medical images, a precision of the registration far below the voxel size can be achieved.

7 Acknowledgments

We want to thank Nicholas Ayache and Mike Brady for stimulating discussions about error handling. Part of that research has been supported by the Esprit Basic Research Action VIVA, and also by a contract with Digital Equipment.

References

[AHB87] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, September 1987.

[Aya91] N. Ayache. *Artificial Vision for Mobile robots - Stereo-vision and Multisensor Perception.* MIT-Press, 1991.

[DW88] H.F. Durrant-Whyte. *Integration, Coordination and Control of Multi-Sensor Robot System.* Kluwer Academic Publishers, 1988.

[FH84] J. Fang and T.S. Huang. Some experiments on estimating the 3-d motion parameters of a rigid body from two consecutive image frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(5):545–554, September 1984.

[HJLZ89] R.M. Haralick, H. Joo, C.N. Lee, X. Zhuang, V.G. Vaidya, and M.B. Kim. Pose estimation from corresponding point data. *IEEE transaction on Systems, Man and Cybernetics*, 19(6):1426–1446, November,December 1989.

[Hor87] B.K.P. Horn. Closed form solutions of absolute orientation using unit quaternions. *Journal of Optical Society of America*, A-4(4):629–642, April 1987.

[Kan93] K. Kanatani. *Geometric Computation for Machine Vision. The Oxford Engineering science series*, Clarendon Press - Oxford, 1993.

[PT95] X. Pennec and J.P. Thirion. *Validation of 3-D Registration Methods based on Points and Frames.* Research Report 2470, INRIA, 1995.

[SB] L.S. Shapiro and M. Brady. Motion and structure clustering using maximum affinity spanning trees. Submitted to ICCV95.

[Sny89] M.A. Snyder. The precision of 3-d parameters in correspondence-based techniques: The case of uniform translational motion in a rigid environment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):523–528, May 1989.

[Thi93] J.P. Thirion. *New Feature Points based on geometric invariants for 3D Image Registration.* Research Report 1901, INRIA, Mai 1993. accepted for publication in IJCV.

[Thi94] J-P Thirion. Extremal points : definition and application to 3d image regist ration. In *IEEE conf. on Computer Vision and Pattern Recognition*, Seattle, June 1994.

[Ume91] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, April 1991.

[ZF92] Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis: a stereo based approach*, chapter 2: Uncertainty Manipulation and Parameter Estimation, pages 9–27. Volume 27 of *Springer series in information science*, Springer Verlag, 1992.

[ZH94] X.H. Zhuang and Y. Huang. Robust 3-d 3-d pose estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):818–824, August 1994.