To appear in Proc. of the 8th European Conference on Computer Vision - ECCV 2004, Prague, May 11-14, 2004.

An Accuracy Certified Augmented Reality System for Therapy Guidance

S. Nicolau[2,1], X. Pennec[1], L. Soler[2], N. Ayache[1]

[1] INRIA Sophia, Epidaure, 2004 Rte des Lucioles, F-06902 Sophia-Antipolis Cedex
{Stephane.Nicolau, Xavier.Pennec, Nicholas.Ayache}@sophia.inria.fr
http://www-sop.inria.fr/epidaure/Epidaure-eng.html

[2] IRCAD-Hopital Civil, Virtual-surg, 1 Place de l'Hopital, 67091 Strasbourg Cedex
{stephane.nicolau, luc.soler}@ircad.u-strasbg.fr

Abstract. Our purpose is to provide an augmented reality system for Radio-Frequency guidance that could superimpose a 3D model of the liver, its vessels and tumors (reconstructed from CT images) on external video images of the patient. In this paper, we point out that clinical usability not only need the best affordable registration accuracy, but also a certification that the required accuracy is met, since clinical conditions change from one intervention to the other. Beginning by addressing accuracy performances, we show that a 3D/2D registration based on radio-opaque fiducials is more adapted to our application constraints than other methods. Then, we outline a lack in their statistical assumptions which leads us to the derivation of a new extended 3D/2D criterion. Careful validation experiments on real data show that an accuracy of 2 mm can be achieved in clinically relevant conditions, and that our new criterion is up to 9% more accurate, while keeping a computation time compatible with real-time at 20 to 40 Hz.

After the fulfillment of our statistical hypotheses, we turn to safety issues. Propagating the data noise through both our criterion and the classical one, we obtain an explicit formulation of the registration error. As the real conditions do not always fit the theory, it is critical to validate our prediction with real data. Thus, we perform a rigorous incremental validation of each assumption using successively: synthetic data, real video images of a precisely known object, and finally real CT and video images of a soft phantom. Results point out that our error prediction is fully valid in our application range. Eventually, we provide an accurate Augmented Reality guidance system that allows the automatic detection of potentially inaccurate guidance.

1 Introduction

The treatment of liver tumors by Radio-Frequency (RF) ablation is a new technique which begins to be widely used [11]. However, the guidance procedure to reach the tumors with the electrode is still made visually with per-operative 2D

cross-sections of the patient using either Ultra-Sound (US) or Computed Tomography (CT) images. Our purpose is to build an augmented reality system that could superimpose reconstructions of the 3D liver and tumors onto video images in order to improve the surgeon's accuracy during the guidance step. According to surgeons, the overall accuracy of such a system has to be less than 5 mm to provide significant help.

In our setup, a CT-scan of the patient is acquired just before the intervention (RF is a radiological act), and an automatic 3D-reconstructions of his skin, his liver and the tumors is performed [2]. Two cameras (jointly calibrated) are viewing the patient's skin from two different points of view. The patient is intubated during the intervention, so the volume of gas in his lungs can be controlled and monitored. Then, it is possible to fix the volume at the same value during a few seconds repetitively and to perform the electrode's manipulation almost in the same volume's condition than the one obtained during the preliminary CT-scan. Balter [1] and Wong [20] indicates that the mean tumor repositioning at exhalation phase in a respiratory-gated radiotherapy context is under 1 mm. Thus, it is reasonable to assume that a rigid registration is sufficient to register accurately the 3D-model extracted from the CT with the 2D video images.

Critical issues for computer-guided therapy systems are accuracy and reliability. Indeed, the surgeon has no other source of information than the augmented reality system during the guidance step: he has to rely fully on it. As many parameters can change from one intervention to the other (angle between the cameras, cameras focal, curvature of the patient abdomen), the accuracy provided can sharply vary. For instance, in a point-based registration context, there can be a factor two on the accuracy when the cameras angle goes from 20^o to 60^o [12]. In accordance with this fact, we cannot afford providing a system without assessing its accuracy during any possible intervention. Consequently, we need to tackle both the system accuracy and the capability to assess its value *before* the intervention. Moreover, every gain in accuracy may be exploited to release some constraints in the system setup (position of the cameras, ergonomics, computation time...).

To answer these requirements, we review in Section 2 the existing registration techniques, and we focus more particularly on 3D/2D points based methods. As their statistical assumptions are not fully satisfied in our application (our 3D point measurements cannot be considered as noise-free), we derive a new criterion that extends the classical one. Experimental results on synthetic and phantom data show that it provides a registration up to 20% more accurate. To be able to quantify online this accuracy, we apply in Section 3 the general theory of error propagation to our new criterion and its standard version. This gives us an *analytical* formulation of the covariance matrix of the seeked transformation. But this is only the first part of the job: we then need to validate this prediction w.r.t. the statistical assumptions used to derive the theoretical formula (small non-linearity of the criterion, perfect calibration, unbiased Gaussian noise on points, etc.). Incremental tests with synthetic data, real cameras, and finally real data of a soft phantom, show that our prediction is reliable for our current

setup, but may require the inclusion of calibration and skin motion errors if it was to become more accurate.

2 A New 3D/2D Point-Based Registration Criterion

This section aims at finding the most accurate registration method for our application.

2.1 Surfacic, Iconic, 3D/3D or 3D/2D Registration?

Surface and iconic registration using mutual information have been used to register the 3D surface of the face to either video images [19] or another 3D surface acquired with a laser range scanner [5]. In both cases, thanks to several highly curved parts on the model (nose, ears, eyes), the reported accuracy was under 5 mm. We believe that in our case, the "cylindrical" shape of the human abdomen is likely to lead to much larger uncertainties along the cranio-caudal axis.

Landmarks 3D/3D or 3D/2D registration can be performed when several precisely located points are visible both in the 3D-model and in the video images. Since the landmarks are really homologous, the geometry of the underlying abdomen surface is not any more a problem. As there are no visible anatomical landmarks in our case, we chose to stick to the patient skin some radio-opaque markers that are currently localized interactively (an automatic segmentation is currently being tested). The matching is performed thanks to epipolar geometry between video points, and using a prediction/verification (alignment) algorithm between video and CT points.

As our system is based on a stereoscopic video acquisition, one could think of using a stereoscopic reconstruction. In our case, the main problem is the possible occlusion of some 2D points in one of the cameras, which would lead to discarding the information provided by this point in the other camera. Moreover, one would need to compute non-isotropic uncertainty of the reconstructed 3D points [8] to optimize a 3D/3D registration criterion fitting well the statistical assumptions. Thus, we believe that it is better to rely on LSQ 3D/2D registration criteria.

The 3D/2D registration problem was largely considered in a wide variety of cases. Briefly, we can classify the different methods in 3 groups: closed-form, linear and non-linear. The two first method classes were proposed in the last decades to find the registration as quickly as possible to fulfill real-time constraints [6, 3, 7]. However they are very sensitive to noise because they assume that data points are exact, contrary to non-linear method. Consequently, non-linear methods provides better accuracy results [10, 18]. As the accuracy is crucial in our application, we think that a LSQ criterion optimization has a definite advantage among the other methods because it can take into account the whole information provided by the data. However, all of the existing methods [4, 9, 15, 10] implicitly consider that 2D points are noisy, but that 3D points of the model to register are exact. In our case, this assumption is definitely questionable, which lead to the development of a new maximum likelihood (ML) criterion generalizing the standard 3D/2D LSQ criterion.

2.2 Maximum Likelihood 3D/2D Registration

Notations Let M_i ($i \in \{1 \cdots N\}$) be the 3D points that represent the exact localization of the radio-opaque fiducials in the CT-scan reference frame and ${m_i}^{(l)}$ be the 2D points that represent its exact position in the images of camera (l). To account for occlusion, we use a binary variable ξ_i^l equal to 1 if M_i is observed in camera (l) and 0 otherwise. We denote by $< \cdot | \cdot >$ the cross-products, by $T \star M$ the action of the rigid transformation T on the 3D point M and by P_l ($1 \leq l \leq S$) the camera's projective functions from 3D to 2D such that $m_i^{(l)} = P^{(l)}(T \star M_i)$. In the following sections, $\hat{A}$ will represent an *estimation* of a perfect data A, and $\tilde{A}$ will represent an *observed measure* of a perfect data A.

Standard Projective Points Correspondences (SPPC) Criterion Assuming that the 3D points are exact ($\tilde{M}_i = M_i$) and that the 2D points only are corrupted by an isotropic Gaussian noise η_i of variance σ_{2D}^2, the probability of measuring the projection of the 3D point M_i at the location ${\tilde{m}_i}^{(l)}$ in image (l), knowing the transformation parameters $\theta = \{T\}$ is given by:

$$p({\tilde{m}_i}^{(l)} \mid \theta) = \frac{1}{2\pi\sigma_{2D}^2} \cdot \exp\left(- \frac{\parallel P^{(l)}(T \star M_i) - {\tilde{m}_i}^{(l)} \parallel^2}{2 \cdot {\sigma_{2D}}^2} \right)$$

Let χ be the data vector regrouping all the measurements, in this case the 2D points ${\tilde{m}_i}^{(l)}$ only. Since the detection of each point is performed independently, the probability of the observed data is $p(\chi \mid \theta) = \prod_{l=1}^{S} \prod_{i=1}^{N} p({\tilde{m}_i}^{(l)} \mid \theta)^{\xi_i^l}$. In this formula, unobserved 2D points (for which $\xi_i^l = 0$) are implicitly taken out of the probability. Now, the *Maximum likelihood* transformation $\hat{\theta}$ maximizes the probability of the observed data, or equivalently, minimizes its negative log:

$$C_{2D}(T) = \sum_{l=1}^{S} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\left\| P^{(l)}(T \star M_i) - \tilde{m}_i^{(l)} \right\|^2}{2 \cdot {\sigma_{2D}}^2} + \left(\sum_{l=1}^{S} \sum_{i=1}^{N} \xi_i^l \right) \cdot \log[2\pi\sigma_{2D}^2] \quad (1)$$

Thus, up to a constant factor, this ML estimation boils down to the classical 3D/2D points LSQ criterion.

Extended Projective Points Correspondences (EPPC) Criterion To introduce a more realistic statistical hypothesis on the 3D data, it is thus safer to consider that we are measuring a noisy version of the exact points: $\tilde{M}_i = M_i + \varepsilon_i \quad \text{with} \quad \varepsilon_i \sim N(0, \sigma_{3D})$.

In this case, the exact location M_i of the 3D points is considered as a parameter, just as the transformation T. In statistics, this is called a *latent or hidden variable*, while it is better known as an *auxiliary variable* in computer vision. Thus, knowing the parameters $\theta = \{T, M_1, \ldots M_N\}$, the probability of measuring respectively a 2D and a 3D point is:

$$p(\tilde{m}_i^{(l)} \mid \theta) = G_{\sigma_{2D}}\left(P^{(l)}(T \star M_i) - \tilde{m}_i^{(l)} \right) \quad \text{and} \quad p(\tilde{M}_i \mid \theta) = G_{\sigma_{3D}}\left(M_i - \tilde{M}_i \right).$$

One important feature of this statistical modeling is that we can safely assume that all 3D and 2D measurements are independent. Thus, we can write the probability of our observation vector $\chi = (\tilde{m}_1^1, ..., \tilde{m}_N^1, ..., \tilde{m}_1^S, ..., \tilde{m}_N^S, \tilde{M}_1, ..., \tilde{M}_N)$ as the product of the above individual probabilities. The ML estimation of the parameters is still given by the minimization of $-\log(p(\chi|\theta))$:

$$C(T, M_1, \cdots M_N) = \sum_{i=1}^{N} \frac{\parallel \tilde{M}_i - M_i \parallel^2}{2 \cdot {\sigma_{3D}}^2} + \sum_{l=1}^{S} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\parallel \tilde{m}_i^{(l)} - P^{(l)}(T \star M_i) \parallel^2}{2 \cdot {\sigma_{2D}}^2} + K$$

where K is a normalization constant depending on σ_{2D} and σ_{3D}. The convergence is insured since we minimize the same positive criterion at each step.

The obvious difference between this criterion and the simple 2D ML is that we now have to solve for the hidden variables (the exact locations M_i) in addition to the previous rigid transformation parameters. An obvious choice to modify the optimization algorithm is to perform an alternated minimization w.r.t. the two groups of variables, starting from a transformation initialization T_0, and an initialization of the M_i with the $\tilde{M}_i$. The algorithm is stopped when the distance between the last two estimates of the transformation become negligible.

Discussion We highlight in [12] that the EPPC can be viewed as a generalization of either the standard criterion (when $\sigma_{3D} \to 0$), or a stereoscopic points reconstruction followed by a 3D/3D registration (if σ_{3D} is largely overestimated). A quantitative study on synthetic data showed that accuracy gain brought by EPPC depends essentially on the angle between the cameras and ratio of 2D and 3D SNR [12]. For instance, with data simulating our clinical conditions, EPPC brings up to 10% gain accuracy if the cameras angle is 50^o and 18% if the angle is 20^o. Finally, as simulation does not take into account calibration errors and possible noise modeling errors, we made a careful validation on real data from a phantom. This showed that a mean accuracy of 2 mm can be reached with a maximum error of 4 mm (obtained when the parameters configuration are not optimal: weak angle between the cameras and/or markers occlusion) and that we can rely on an accuracy gain of 9% with computations time that can still fulfill real-time constraints.

3 Theoretical Uncertainty and Prediction Validation

Now that we have provided a criterion that perfectly fulfills the statistical conditions of our application, we still face the problem of the varying accuracy w.r.t. the various system parameters. In order to propose a safe product to radiologists, we should provide a statistical study that would give the mean *Target Registration Error* (TRE) w.r.t. the number of markers, the angle between the cameras, the focus, and the relative position of the target w.r.t. the markers. This is the equivalent of the direction for use and the secondary effects list mandatory for all proposed drugs in the therapeutic field, and the reliability and accuracy tables of robotics tools: these tables give a usability range to assess under which condition a particular feature (for example accuracy) could be reached.

As increasing the number of experiments is very expensive and time-consuming, it is almost infeasible to *measure* the accuracy provided for each experimental condition. Moreover, as we want a real-time system, the conditions may change during the operation (e.g. markers can be occluded by the radiologist), and the accuracy assessment has to be constantly updated to avoid a potentially dangerous gesture. Consequently, we think that predicting the TRE by studying the theoretical noise propagation is the best way to ensure the safety of our system.

3.1 Uncertainty Propagation through SPPC and EPPC Criteria

In the sequel, we firstly remind the general covariance propagation theory through a criterion. Then, following the methodological framework introduced in [14, 13], we present for SPPC and EPPC analytical formulations of the transformation covariance matrix.

General Theory of Error Propagation Let the criterion $C(\chi, \theta)$ be a smooth function of the data vector χ and the parameters θ. We are looking for the optimal parameter vector $\hat{\theta} = \arg\min_\theta(C(\chi, \theta))$. A local minima is reached and well defined if $\Phi(\chi, \theta) = \left(\frac{\partial C}{\partial \theta}(\chi, \theta)\right)^\top = 0$ and $H = \frac{\partial^2 C}{\partial \theta^2}(\chi, \theta)$ is positive definite. The function Φ defines $\hat{\theta}$ as an implicit function of χ. A Taylor expansion gives:

$$\Phi(\chi + \delta\chi, \theta + \delta\theta) = \Phi(\chi, \theta) + \frac{\partial \Phi}{\partial \chi} \cdot \delta\chi + \frac{\partial \Phi}{\partial \theta} \cdot \delta\theta + O(\delta\chi^2, \delta\theta^2)$$

which means that around an optimum $\hat{\theta}$ we have:

$$\hat{\theta}(\chi + \delta\chi) = \hat{\theta}(\chi) - \left(\frac{\partial \Phi}{\partial \theta}\right)^{(-1)} \cdot \frac{\partial \Phi}{\partial \chi} \cdot \delta\chi + O(\delta\chi^2)$$

Thus, if χ is a random vector of mean $\bar{\chi}$ and covariance $\Sigma_{\chi\chi}$, the optimal vector $\hat{\theta}$ is (up to the second order), a random vector with mean $\bar{\theta} = \arg\min_\theta(C(\bar{\chi}, \theta))$ and covariance $\Sigma_{\theta\theta} = H^{-1}\left(\frac{\partial \Phi}{\partial \chi}\right)\Sigma_{\chi\chi}\left(\frac{\partial \Phi}{\partial \chi}\right)^\top H^{-1}$. Thus, to propagate the covariance matrix from the data to the parameters optimizing the criterion, we need to compute $H = \frac{\partial^2 C(\chi,\theta)}{\partial \theta^2}$, $J_{\Phi_\chi} = \frac{\partial^2 C(\chi,\theta)}{\partial \chi \partial \theta}$ and $\Gamma = J_{\Phi_\chi} \cdot \Sigma_{\chi\chi} \cdot J_{\Phi_\chi}^\top$.

SPPC Transformation Covariance Our analytical analysis needs the block-decomposition of the 3×4 projection matrix 3 as shown below:

$$P^{(l)} = \left[\begin{array}{c|c} Q^{(l)}_{2\times 3} & b^{(l)}_{2\times 1} \\ \hline C^T_{(l)} & 1 \end{array}\right] \qquad \text{so that} \qquad m_i^{(l)} = P^{(l)}(T \star M_i) = \frac{Q^{(l)} \cdot (T \star M_i) + b^{(l)}}{1 + C^T_{(l)} \cdot (T \star M_i)}$$

The second order derivatives H and $J_{\Phi_\chi}^\top$ are computed using the chain rule, and after some calculations, the uncertainty of the transformation may be summarized as $\Sigma_{TT} = H^{-1} \cdot \Gamma \cdot H^{-1}$ with

$$\Gamma = \sum_{i=1}^N {D_i}^T \left(\sigma_{3D}^2 \cdot K_i \cdot K_i + L_i\right) \cdot D_i \qquad \text{and} \qquad H = \sum_{i=1}^N {D_i}^T \cdot K_i \cdot D_i$$

where $D_i = \frac{\partial (T \star M_i)}{\partial T}$, $L_i = \sum_{l=1}^{S} \xi_i^l \cdot \frac{(Q - m_i^{(l)} \cdot C_{(l)}^T)^T \cdot (Q - m_i^{(l)} \cdot C_{(l)}^T)}{\sigma_{2D}^2 \cdot (1 + <C_{(l)} | T \star M_i>)^2}$ and

$K_i = L_i - \sum_{l=1}^{S} \xi_i^l \cdot \frac{C_{(l)} \cdot (m_i^{(l)} - \tilde{m_i}^{(l)})^T \cdot (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^T) + (Q^{(l)} - m_i^{(l)} \cdot C_{(l)}^T)^T \cdot (m_i^{(l)} - \tilde{m_i}^{(l)}) \cdot C_{(l)}^T}{\sigma_{2D}^2 \cdot (1 + <C_{(l)} | T \star M_i>)^2}$

EPPC Transformation Covariance For this case the calculations are not usual because the vector of sought parameters is $\theta = (T, M_1 \cdots M_N)$ so that:

$$\delta\theta = \begin{bmatrix} \delta T \\ \delta M \end{bmatrix} = - \left(\frac{\partial \Phi}{\partial \theta} \right)^{-1} \cdot \frac{\partial \Phi}{\partial \chi} . \delta\chi = - \begin{bmatrix} \frac{\partial \Phi_T}{\partial T} & \frac{\partial \Phi_T}{\partial M} \\ \frac{\partial \Phi_M}{\partial T} & \frac{\partial \Phi_M}{\partial M} \end{bmatrix}^{(-1)} \cdot \begin{bmatrix} \frac{\partial \Phi_T}{\partial \chi} \\ \frac{\partial \Phi_M}{\partial \chi} \end{bmatrix} . \delta\chi$$

Since we only focus on the covariance of the transformation T alone, we need to extract Σ_{TT} from $\Sigma_{\theta\theta} = \begin{bmatrix} \Sigma_{TT} & \Sigma_{TM} \\ \Sigma_{MT} & \Sigma_{MM} \end{bmatrix}$. This is done using a block matrix inversion, and after long calculations, we end up with $\Sigma_{TT} = H^{-1} \cdot \Omega \cdot H^{-1}$, where:

$$\Omega = \sum_{i=1}^{N} {D_i}^T \ \left(\sigma_{3D}^2 Id + K_i^{-1} \right)^{-1} \cdot \left(\sigma_{3D}^2 Id + K_i^{-1} \cdot L_i \cdot K_i^{-1} \right) \cdot \left(\sigma_{3D}^2 Id + K_i^{-1} \right)^{-1} \cdot D_i$$
$$H = \sum_{i=1}^{N} {D_i}^T \cdot \left(\sigma_{3D}^2 Id + K_i^{-1} \right)^{-1} \cdot D_i$$

One can check that for the limit case where $\sigma_{3D} = 0$, the transformation uncertainty given by Σ_{TT} is equal for both criteria.

Target Registration Error (TRE) Finally, to obtain the final covariance matrix on a target point C_i after registration, we simply have to propagate the uncertainty through the transformation action: $\Sigma_{T \star C_i} = \frac{\partial (T \star C_i)}{\partial T} \cdot \Sigma_{TT} \cdot \frac{\partial (T \star C_i)}{\partial T}^T$.

3.2 Validation of the Prediction

With the previous formulas, we are able to predict the accuracy of the transformation after the convergence of the algorithm of Section 2. But this is only one part of the job: we now have to validate the statistical assumptions used to derive the theoretical formula (small non-linearity of the criterion, perfect calibration, unbiased Gaussian noise on points, etc.). The goal of this section is to verify incrementally that these assumptions hold within our application domain. This will be done using synthetic data (for the non-linearities of the criterion), real video images of a precisely defined 3D object (for camera calibration and distortions), and finally real CT and video images of a soft phantom of the abdomen (for noise assumptions on point measurements).

Synthetic Data Experiments are realized with two synthetic cameras jointly calibrated with a uniform angle from 5^o to 120^o, and focusing on 7 to 25 points M_i randomly distributed in a volume of about $30 \times 30 \times 30$ cm^3. The cameras

are located at a distance of the object of 20 to 50 times the focal length. We add to the 2D and 3D points a Gaussian noise with σ varying from 0.5 to 4.0 (which corresponds to a SNR of 60 dB to 90 dB[3]). The registration error is evaluated using control points C_i to assess a TRE instead of a *Fiducial Localization Error* (FLE).

Since each experiment is different, we need to evaluate the relative fit of the *Predicted TRE* (PTRE) vs. the *Experimental* TRE (ETRE) to quantitatively measure the quality of the uncertainty prediction. Due to the significant anisotropy, we did not use the basic ratio $ETRE^2/PTRE^2$, but rather the validation index [14], which weights the observed error vector with the inverse of its predicted covariance matrix to yield a Mahalanobis distance μ^2. Assuming a Gaussian error on test points after registration, this validation index should follows a χ^2_3 law. Repeating this experiment with many different "parameters" configurations, we can verify that μ^2 is actually χ^2_3 distributed using the Kolmogorov-Smirnov (K-S) test [16]. We also verify that the empirical mean and variance matches the theoretical ones (resp. 3 and 6 for a χ^2_3 distribution).

Table 1 summarizes the statistics obtained for 20000 registrations where all the parameters randomly vary as previously described. The values obtained for both the validation index and the KS-test fully validate the reliability of the transformation's accuracy prediction.

	Mean μ^2 (3.0)	Var μ^2 (6.0)	KS-test ($p > 0.01$)
SPPC	3.020	6.28	0.353
EPPC	3.016	6.18	0.647

Table 1. Validation of the uncertainty prediction with 20000 registrations.

Real Calibration and Synthetic Noise The perfect validation of our accuracy prediction on synthetic data does not take into account possible calibration errors of the cam eras and excludes likely distortions from the pinhole model. The goal of this experiment is to address the validity of these assumptions using a real video system. We used a 54 points calibration grid that allows for a very accurate detection of the points ($\sigma_{3D} < 0.1$ mm, $\sigma_{2D} < 0.2$ pixel). Such an accuracy is obviously far below the current detection of real markers positions ($\sigma_{2D} \simeq 2$ pixel, $\sigma_{3D} \simeq 1$ mm). To simulate the range of variability of our application, we still add a Gaussian noise on the collected data points.

Ideally, the *Experimental* TRE should be assessed by comparing each registration result with a gold-standard that relates both the CT and the camera coordinate systems to the same physical space, using an external and highly accurate apparatus. As such a system is not available, we adapted the registration loops protocol introduced in [13, 17], that enables to measure the TRE error for a given set of test points.

[3] $SNR_{dB} = 10 \log_{10}(\frac{\sigma_s}{\sigma_n})$ where σ_s (resp. σ_n) is the variance of the signal (resp. noise).

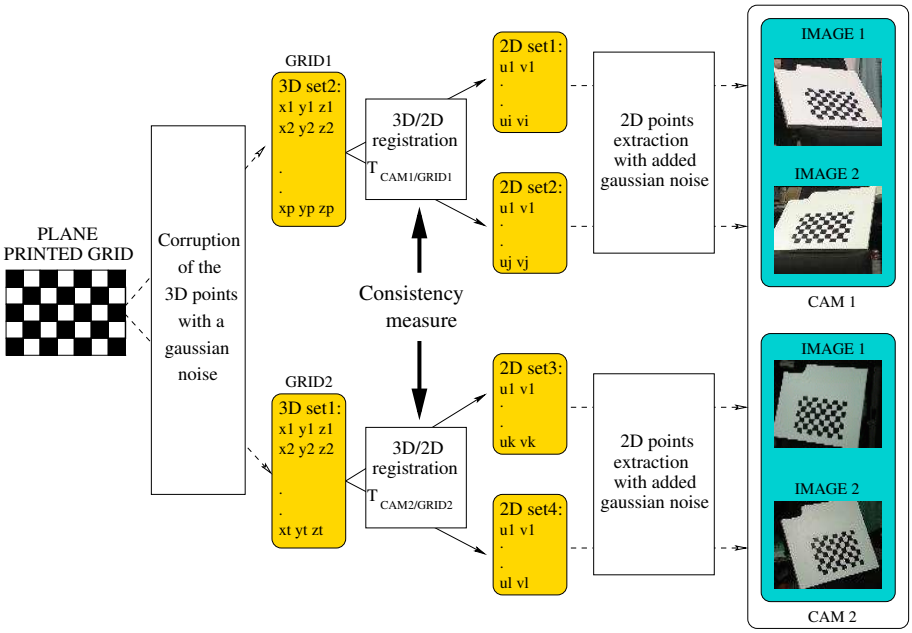

Fig. 1. Registration loops used to estimated the registration consistency: a test point C chosen at a certain distance of the printed grid (typically 20 cm) is transformed into the CAM1 coordinate system using a first 3D/2D registration T_1, then back into the grid coordinate system using a second 3D/2D registration T_2 provided by the other couple of cameras (the coordinate system of CAM1 and CAM2 are identical since cameras are jointly calibrated). If all transformations were exact, we would obtain the same position for the test point. Of course, since the transformations are not perfect, we measure an error which variance ${\sigma^2}_{loop} = 2{\sigma^2}_{CAM/GRID}$ corresponds to a TRE. In fact, to take into account anisotropies we compute a covariance matrix and a statistical Mahalanobis distance μ^2 between C and $T_1 \star T_2^{-1} \star C$.

The principle is to acquire several couples of 2D images with jointly calibrated cameras so that we can compare *independent* 3D/2D registration of the same object (different 2D and 3D images) using a statistical Mahalanobis distance μ^2. A typical loop, sketched in Fig. 1, described the method to get a μ^2-value. This experiment providing only one error measurement, we still need to repeat it with different datasets to obtain statistically significant measures. In order to take into account possible calibration error and/or bias, it is necessary to change the cameras calibrations and positions, and not only to move the object in the physical space. Likewise, to decorrelate the two 3D/2D transformations, we need to use two differently noised 3D data sets. Indeed, when using the same set of 3D points to register the 2D points, the error on 3D points similarly affects both transformations, and the variability of the 3D points extraction (and any possible bias) is hidden.

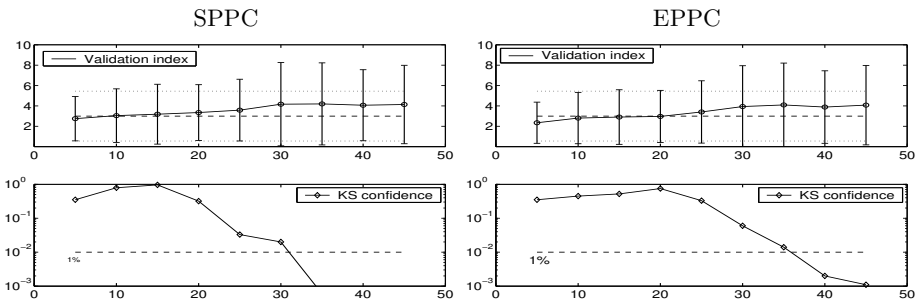

Fig. 2. Validation of the uncertainty prediction on the calibration grid w.r.t. the number of points used for the registration. Top: mean and standard deviation of the validation index. Bottom: KS confidence. Higher scores are more confident.

Finally, varying each set of parameters (different configuration of our four cameras, different positions/orientations of the calibration grid), we got 144 μ^2-values. The cameras were placed 10^o to 60^o apart, at a distance of the object of 25 to 30 times the focal length. Figures 2 shows the mean, standard deviation and K-S test value of the validation index w.r.t. the number of points used (randomly chosen among the 54 available). One can see that the prediction is correct up to 40 points (which spans our range of application). This critical value is due to the progressive reduction of the registration error that finally meets the ignored calibration error (about 0.5 mm). Likewise, we observed the same behavior when the feature noise becomes too small (σ_{3D} and σ_{2D} below 0.7).

Real Data (Phantom) To test the last assumption of our prediction (unbiased Gaussian noise), we now turn to a validation experiment on real 3D and 2D measurements of a plastic phantom (designed in [12]), on which are stick about 40 radio-opaque markers (see the incrusted top left image in Fig. 4). The set up is almost the same as for the previous calibration grid (for further details see [12]). However, target points C_i are now randomly chosen within the phantom liver, and markers in the CT and on the video images are interactively localized.

The markers used were randomly chosen among the 40 available, and we obtained 80 μ^2-values for each experiment. As we experimentally observed that there was a consistent but non-rigid motion of the soft skin (about 1mm), we chose $\sigma_{3D} \simeq 2.0$ mm (instead of 1 mm) to take into account this additional uncertainty. Figure 3 presents the mean and variance of μ^2 w.r.t. the number of points. Firstly, we notice that the mean value slowly increases with the number of points. This can be explained by the biases introduced by the calibration error and the correlated motion of the markers on the skin. Indeed, the measured accuracy figures do not converge to 0 mm with a large number of points but rather towards 1 mm, which corresponds to the motion of the skin.

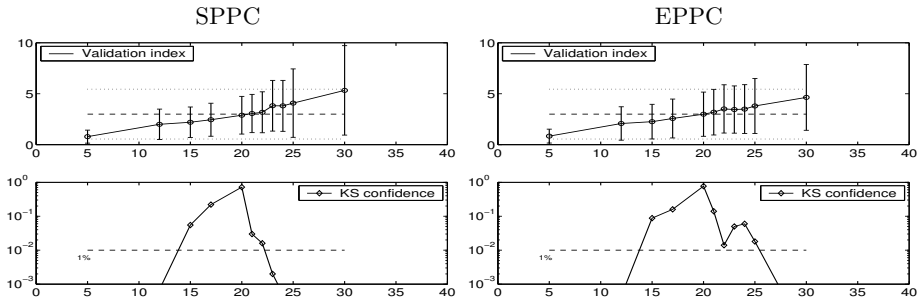

Fig. 3. Validation of the uncertainty prediction on the phantom w.r.t. the number of points used for the registration. Top: mean and standard deviation of the validation index. Bottom: KS confidence. Higher scores are more confident.

Nevertheless, it appears that the prediction is well validated for a range of 15 to 25 points. As $\bar{\mu}^2$ can be interpreted as a relative error or the error prediction (see [14]), Fig. 3 shows that we over-estimate the mean TRE by a factor 1.7 for a small number of points ($\bar{\mu}^2 \simeq 1$), and that we under-estimate it by a factor of 1.3 for more than 25 points ($\bar{\mu}^2 \simeq 5$). For our application, in which the number of visible points should not exceed 20, this means that we predict correctly the amplitude of the error on the transformation. In the worst case, we over-estimate it, which can be considered as a good safety measure. One can visually assess the validation of our prediction error on one case among the 160 registrations we performed (Fig.4).

Fig. 4. Left image: the top image shows the phantom with radio-opaque markers on its skin. The main image shows the phantom without its skin and we can see the radio-opaque markers on the fake liver. Right image: we superimpose the reconstructions of the fiducials whose predicted accuracy is around 2 mm. One can visually assess the quality of the registration.

4 Conclusion

We devised in this paper an augmented reality system for Radio-Frequency ablation guidance based on a new 3D/2D registration criterion with a validated error prediction. We argue the necessity to provide not only the best affordable registration accuracy but also an accurate assessment of the TRE for safety consideration.

To reach the best accuracy performances, we firstly derived a new 3D/2D Maximum Likelihood registration criterion (EPPC) based on better adapted statistical hypotheses than the classical 3D/2D least-square registration criterion (SPPC). Experiments on real data showed that EPPC provides an accuracy of about 2mm within the liver, which fits the initial requirements of less than 5mm. Moreover, EPPC is up to 9% more accurate than SPPC with a refreshment rate that can reach real-time constraints. We underline an alternative interpretation of this gain: we can typically reach the same accuracy with 20 markers for EPPC where 24 are needed for SPPC. As we face possibilities of markers occlusion because of the surgeon's hand and cumbersomeness constraints on the placement of the markers, this gain should not be taken with the light one. In addition, as clinical conditions do not allow a free camera positioning, we could meet situation where an angle between the cameras could decrease below 20^o, which would mean an accuracy gain of 18%.

In order to assess the system accuracy for all configurations, we propose in a second step a theoretical propagation of the target covariance through SPPC and EPPC w.r.t the experimental configuration parameters. To verify the validity of all the assumptions of that method, we conducted a careful validation study that assess in turn the range of validity of each hypothesis. We firstly verified that non-linearities in the criterion and calibration error are negligible. Then, we use a realistic phantom with a soft and deformable skin to validate the prediction in the range of our application (i.e. for 15 and 25 markers). This study confirmed that we correctly predict the registration error, with a slight over-estimation if too much markers are occluded, which is a good safety rule.

To reach the clinical usability, the whole system still has to be validated on real patients. We are currently conducting experiments (using repeated CT scans at the same point of the breathing cycle) to certify that the motion of the internal structures due to the monitored breathing of the patient cannot bias our accuracy prediction. Preliminary results indicates that this motion is of the order of 1 mm, which is in accordance with the motions we experienced because of the phantom soft skin. Thus, we are pretty confident that our registration error prediction will work properly in the final system. Last but not least, it is possible to estimate broadly the TRE *before* scanning the patient, by using the stereoscopic reconstruction of the markers instead of their positions in the scanner. This will allow a better control of the external conditions (number of markers, angle between the cameras) and the optimization of the intervention preparation.

References

1. J.M. Balter, K.L. Lam, C.J. McGinn, T.S. Lawrence, and R.K. Ten Haken. Improvement of CT-based treatment-planning models of abdominals targets using static exhale imaging. *Int. J. Radiation Oncology Biol. Phys.*, 41(4):939–943, 1998.
2. L. Soler et al. Fully automatic anatomical, pathological, and functional segmentation from CT-scans for hepatic surgery. *Computer Aided Surgery*, 6(3), 2001.
3. M. Dhome et al. Determination of the attitude of 3D objects from a single perspective view. *IEEE Trans. on PAMI*, 11(12):1265–1278, December 1989.
4. R. Haralick et al. Pose estimation from corresponding point data. *IEEE Trans. on Systems, Man. and Cybernetics*, 19(06):1426–1446, December 1989.
5. W. Grimson et al. An automatic registration method for frameless stereotaxy, image-guided surgery and enhanced reality visualization. *IEEE TMI*, 15(2):129–140, April 1996.
6. M. Fischler and R. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Com. of the ACM*, 24(6):381–395, June 1981.
7. S. Ganapathy. Decomposition of transformation matrices for robot vision. *Pattern Recognition Letters*, 2(6):401–412, December 1984.
8. E. Grossmann and J. Santos-Victor. Uncertainty analysis of 3D reconstruction from uncalibrated views. *Image and Vision Computing*, 18:685–696, 2000.
9. Y. Hel-Or and M. Werman. Pose estimation by fusing noisy data of different dimensions. *IEEE Trans. on PAMI*, 17(2):195–201, February 1995.
10. C. Lu, G.D Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Trans. on PAMI*, 22(6):610–622, 2000.
11. J.F. McGahan and G.D. Dodd III. Radiofrequency ablation of the liver: Current status. *American Journal of Roentgenology*, 176(1):3–16, 2001.
12. S. Nicolau, X. Pennec, L. Soler, and N. Ayache. Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by augmented reality. In *IS4TM'03*, volume 2673 of *LNCS*, pages 270–283, France, 2003. Springer-Verlag.
13. X. Pennec, C.R.G. Guttmann, and J.-P. Thirion. Feature-based registration of medical images: Estimation and validation of the pose accuracy. In *MICCAI'98*, LNCS 1496, pages 1107–1114, October 1998.
14. X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3D registration methods based on points and frames. *IJCV*, 25(3):203–229, 1997.
15. T. Phong, R. Horaud, and P. Tao. Object pose from 2-D to 3-D points and line correspondences. *IJCV*, 15:225–243, 1995.
16. W.H. Press, B.P. Flannery, S.A Teukolsky, and W.T. Vetterling. *Numerical Recipices in C*. Cambridge Univ. Press, 1991.
17. A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE TMI*, 20(10):1038–1049, 2001.
18. J. Salvi, X. Armangu, and J. Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, 35(7):1617–1635, 2002.
19. P. Viola and W.M. Wells. Alignment by maximization of mutual information. *International Journal of Computer Vision*, 24(2):137–154, 1997.
20. J. Wong, M. Sharpe, D. Jaffray, V. Kini, J. Robertson, J. Stromberg, and A. Martinez. The use of active breathing control (ABC) to reduce margin for breathing motion. *Int. J. Radiation Oncology Biol. Phys.*, 44(4):911–919, 1999.