Pre-print of an article to appear in the Journal of Computer Animation and Virtual Worlds in 2004. Nadia Magnenat-Thalmann and Daniel Thalmann editors. Wiley publisher.

An Augmented Reality System to Guide Radio-Frequency Tumor Ablation

S. Nicolau[1,2], A. Garcia[1], X. Pennec[2], L. Soler[1], N. Ayache[2].
[1] IRCAD, 1 Place de l'Hôpital, 67091, Strasbourg, France.
[2] INRIA, Projet Epidaure, 2004 Route des Lucioles, 06902, Sophia Antipolis, France.
Phone number: +33 (0) 3 88 11 90 95
Fax number: +33 (0) 3 88 11 91 92
{stephane.nicolau, luc.soler}@ircad.u-strasbg.fr
Keywords : augmented reality, radio-frequency ablation, 3D/2D registration, needle tracking.

Abstract

Radio-frequency ablation is a difficult operative task that consists in a precise needle positioning in the center of the pathology. This article presents an Augmented Reality system for hepatic therapy guidance that superimposes in real-time 3D reconstructions (from CT acquisition) and a virtual model of the needle on a patient's external views. The superimposition of reconstructed models is performed with a 3D/2D registration based on radio-opaque markers stuck on the patient skin. The characteristics of the problem (accuracy, robustness and time processing) led us to develop automatic procedures to extract and match the markers and to track the needle in real-time. Experimental study confirmed that our algorithms are robust and reliable. Preliminary experiments conducted on a human abdomen phantom showed that our system is highly accurate (needle positioning error within 3mm) and enables the surgeon to reach a target in less than 1 minute on average. Our next step will be to perform an in-vivo evaluation.

Introduction

The treatment of liver tumors by radio-frequency ablation (RFA) is an evolving technology using coagulative necrosis to treat patients with unresectable primary or metastatic hepatic cancers [1]. The guidance procedure to reach tumors with the needle is usually performed visually using intraoperative two dimensional (2D) cross-sections of the patient obtained with either ultrasound (US) or computed tomography (CT) acquisition. Because of the difficulty to locate the tumor's center in three dimensions (3D), the needle positioning is not very accurate and the targeting procedure is very time consuming, since it requires many trials.

Real-time superimposition of images reconstructed in 3D from CT acquisition onto a real patient, so-called Augmented Reality (AR), may improve the accuracy and decrease complications in interventions such as RFA needle placement. Such AR guidance systems are routinely used in the field of neurosurgery and orthopedic surgery [2][3][4][5][6][7][8]. However, there are few applications on abdominal and thoracic zones. Mourgues [9] superimposes coronary arteries on endoscopic images with an accuracy about 5 pixels and Langø [10] provides preliminary results in laparoscopy with an accuracy of about 1 cm. Our purpose is to build a guidance system for the needle positioning that superimposes 3D reconstructions of the liver and its tumors onto video images of the patient abdomen. To provide a significant help to surgeons, the overall superimposition error has to be less than 5 mm and the computation time of the data processing has to be under 10 minutes.

In this article, we present the technical part of our AR system. Twenty five radio-opaque markers are stuck on the skin of the human abdomen (cf. phantom in Figure 1a), which is CT-scanned afterwards (slice thickness of 1 mm). A 3D segmentation, followed by a reconstruction (skin, liver and markers) is performed with a specific software [11] (cf. Figures 1b and 1c). Two jointly calibrated cameras [12] are oriented toward the patient from two different points of view. To superimpose the 3D reconstruction in the video images, we performed a 3D/2D points registration between the markers localized on the reconstruction and those visible in the video images. Since an interactive procedure to localize and match the markers in the CT-scan and video images takes several minutes, we need to realize these tasks automatically. Firstly, we describe the algorithms we developed to extract and match automatically the markers from the CT-scan and video images. Secondly, we present the criterion used to find the rigid transformation that relates the CT-scan and the camera reference frame. Then, we explain how we track the needle in real-time to superimpose its virtual model. Finally, we show our preliminary experiments made on our abdominal phantom.

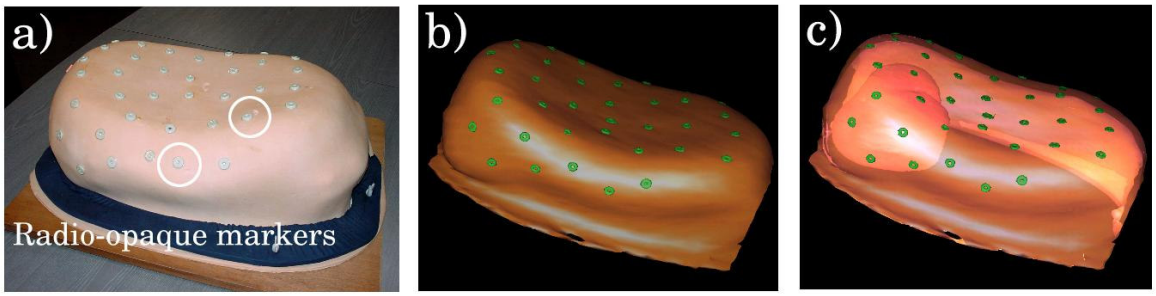

Figure 1: (a) Video image of the phantom on which 25 radio-opaque markers were stuck. (b) 3D reconstruction of the phantom from CT slices. (c) Organ view in transparency. The liver appears under the skin.

2D Marker extraction from video images

The principle of the marker localization in the video images is based on a color analysis and the assumption that the skin covers the main surface area in the video images. In a real intraoperative image, the operative field may modify this assumption. However, it has a homogeneous green color and will cover the border part of the image. Therefore it will be easily identified.

The first step is to find the skin in the image. Working on HSV images, the mean color of the skin $\mathbf{MC_{skin}}$ is given, in the hue histogram, by the mode position with the highest integral area, and not necessary the highest value. Indeed, a small region, more homogeneous than the skin, can provide a mode with a lower standard deviation but with a higher maximum value. Then, a color thresholding is computed around the skin hue. Considering that the skin hue distribution is approximately Gaussian, a good choice for the threshold values $\mathbf{T_{min}}$ and $\mathbf{T_{max}}$ is:

$$T_{\min} = MC_{skin} - 3 \cdot \sigma_{skin} \quad \text{and} \quad T_{\max} = MC_{skin} + 3 \cdot \sigma_{skin}$$

where σ_{skin} is the standard deviation of the skin mode. To evaluate σ_{skin}, we can compute the full width at half max (FWHM) of the skin hue distribution, which is theoretically equal to $2.3548\sigma_{skin}$ under the Gaussian assumption. However, as the histogram is noisy and not totally smooth, it is preferable to evaluate σ_{skin} from a smoothed version of the original hue histogram. The optimal histogram is defined as the iteratively smoothed original histogram with a Gaussian filter until its number of modes becomes stable. We call σ_{opt} the standard deviation of the smoothed hue distribution of the skin and $\mathbf{Max}_{\mathbf{opt}}$ (resp. $\mathbf{Max}_{\mathbf{skin}}$) its maximum value (resp. the maximum value of the original skin hue distribution). Since smoothing does not modify the integral area of the distribution mode, σ_{skin} is related to σ_{opt} by:

$$\sigma_{skin} \cdot Max_{skin} = \sigma_{opt} \cdot Max_{opt}$$

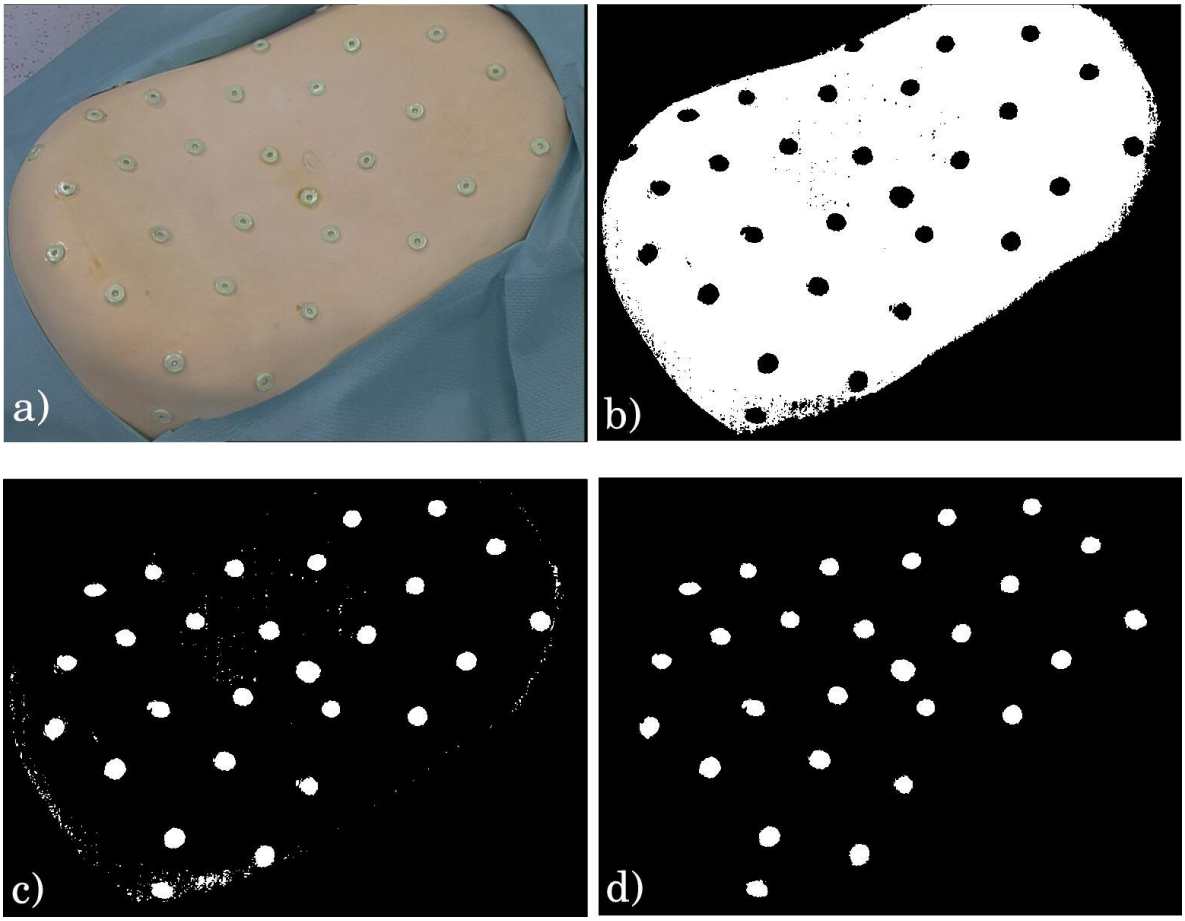

Figure 2: (a) Original image of the phantom. (b) Result of the color thresholding and largest component extraction. (c) The potential markers are all the components in white. (d) Final extraction of markers into the detected skin. The real markers are characterized among the components with respect to their surface and shape.

After this step, as we suppose that the skin covers the largest area in the image, we extract the largest connected component. Figure 2b) shows the results after the color thresholding and the largest component extraction. Then, since the color of markers is very different from skin color (the markers are green), it can be considered that all the connected components entirely enclosed in the detected skin are potential markers. To characterize the real markers among all these components (cf. Figure 2c), we first remove those whose surface is empirically too small to be a marker (under 10 pixels of surface). Then, we compute the median of component surfaces S_{median} and we get rid of the components whose surface is not in the interval $[0.5 \cdot S_{median} ; 2.0 \cdot S_{median}]$. Finally, we discard the components whose inertial moment is too small (characterizing the flat components).

The remaining components (cf. Figure 2d) are considered as markers and their barycenter is computed. To evaluate the performance of this algorithm, we conducted an experimental study on 10 subjects. We stuck around 20 markers on each person and we took pictures of their abdomen in 4 different positions with 2 cameras. From these 80 images, the mean computation time was around 1 second and we had 5% of false negatives and 2% of false positives whose location were never on the skin. Visual inspection indicates that the detection accuracy is around 2 pixels on average. This 2D extraction algorithm is then fast and robust enough to be easily used in a matching algorithm involving 2D and 3D points.

3D Marker extraction from the CT-scan

In order to localize the markers in CT-scan images (cf. Figure 3a), we essentially use mathematical morphological tools. In a first step, we threshold the 3D image to extract both patient and radio-opaque markers. CT-scans being calibrated (Hounsfield Units), a CT independent threshold value can be chosen (it is typically over the air density: -800 HU for the phantom and -300 for a human). In a second step, since some artefacts can appear (like the CT table), we extract the largest connected component from the binary images to get only the patient, and we fill the internal cavity out (cf. Figure 3b). Then, in order to erase the markers, we perform an opening with an adapted structuring element (a sphere with a radius 3 times larger than the marker diameter) (cf. Figure 3c). Eventually, a simple subtraction between the opened image and its previous version provides the markers.

The opening erases the markers but smoothes the surface as well, therefore, residual pieces of surface can remain (cf. Figure 3d). We get rid of them by opening the components found with a small structuring element and by characterizing them with their volume and shape (the same procedure as for video markers). Since the radio-opaque markers are designed to give a high response in CT images, we compute their intensity weighted barycenter **G** to take into account partial volume effects:

$$G = \frac{\sum_{P \in Comp} I(P) \cdot P}{\sum_{P \in Comp} I(P)}$$

where **P** is a 3D position of a voxel in the component and **I(P)** is the voxel intensity. An evaluation performed on 6 CT-scans of the phantom and 3 of a pig, on which were stuck about 25 markers, showed that there was less than 5% of false negatives and 2.3% of false

positives. Most of them corresponded to the pig nipples or to pipes lying along its body. Their presence had no incidence since they were easily identified and discarded during the robust matching step.

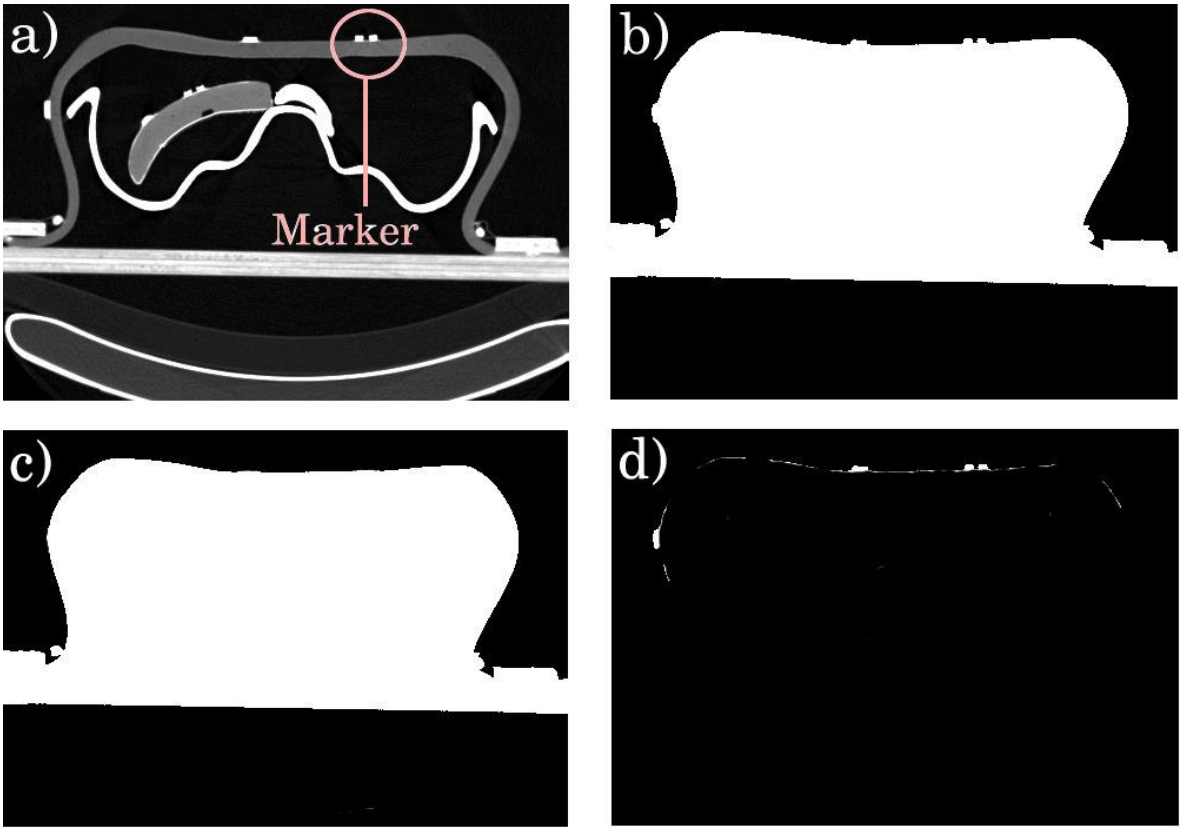

Figure 3: a) Original CT-slice. A radio-opaque marker is circled. b) Original image after the thresholding and the filling. c) Opening of the largest component. Markers have disappeared. d) Subtraction of image (c) to image (b) provides the markers.

Video marker matching

The matching between the 2D markers extracted from the two video images is made with epipolar geometry constraints. As our cameras are jointly calibrated, it is possible to define a relationship between a point in one image and its correspondent in the other one. If **m** is a point in the image **P**, its correspondent **m'** in the image **P'** lies on a segment whose location depends on the optical center **C** and **C'** of the cameras, the position of **m** and the range of depth to which the phantom belongs (cf. Figure 4). Because of imperfect camera calibration and noise, **m'** is not perfectly on the predicted segment, therefore we select the points that are very close to it. If there is only one point, we consider the match as reliable,

otherwise it is ambiguous (see Figure 5). In that case, it is discarded at this step (these markers will be reconsidered in the next step).

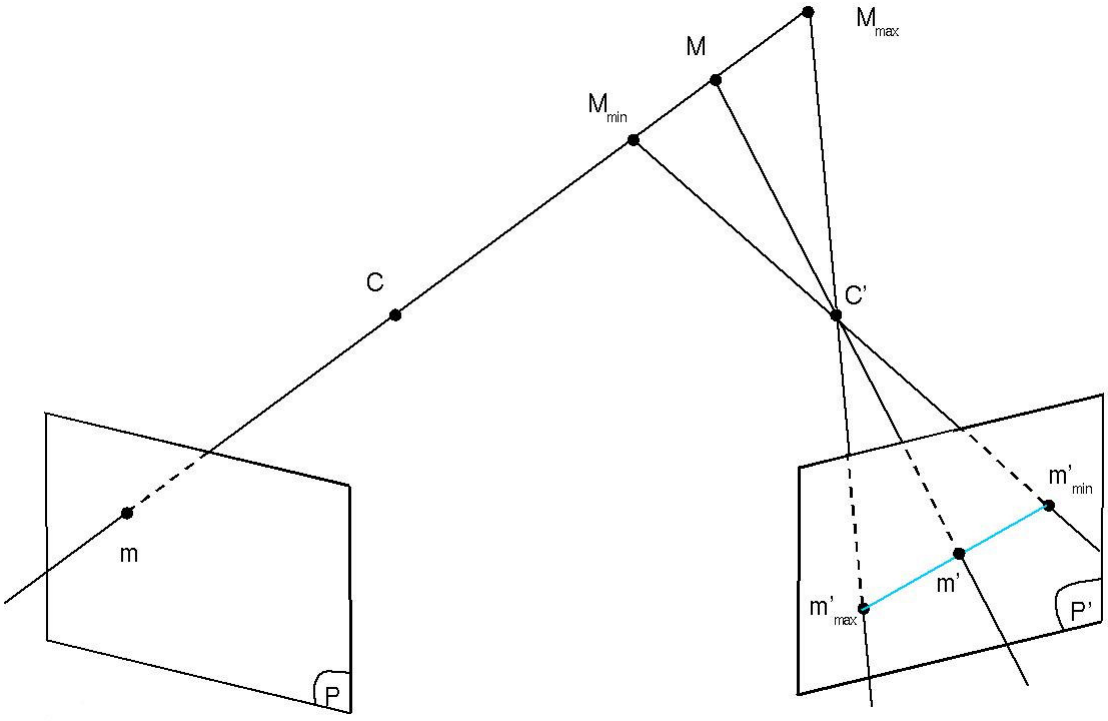

Figure 4: Illustration of the epipolar geometry. Given a 2D marker **m** in the video image **P**, we seek its correspondent **m'** in the video image **P'**. The 3D marker **M**, whose projection on **P** is **m,** lies on the line (**mC**). Since we know the range of depth to which the patient belongs, we can define a segment [M_{min} M_{max}] on (**mC**) that contains **M**. By projecting this segment on the video image **P'**, we build the segment [m'_{min} m'_{max}] on which **m'** lies.

Finally the matched markers are reconstructed in 3D in the camera reference frame. If erroneous 3D reconstructions due to false point matches appear, they are discarded by the subsequent video/CT matching procedure designed to be robust in the presence of such outliers. Experiments realized with 210 image pairs of the phantom showed that on 25 potential matches, an average of 15 were correct, 8.7 were ambiguous and 1.3 were wrong.

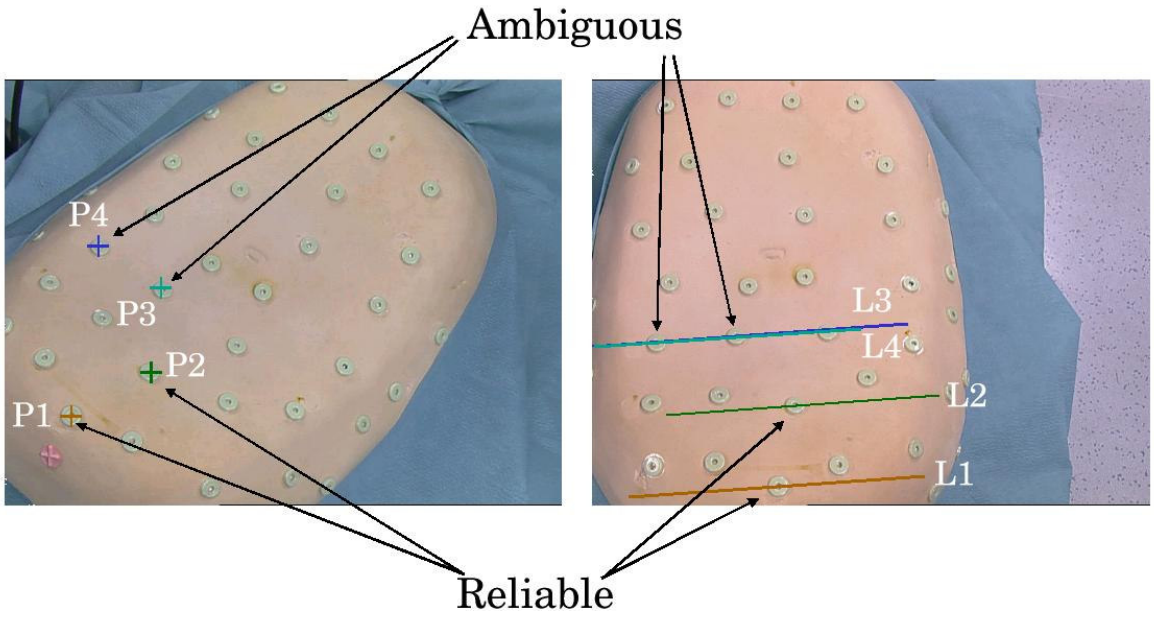

Figure 5: Example of epipolar matching on a video image pair. On the right image are drawn the epipolar segments that correspond to the outlined markers in the left image. One can see that only one marker lies on L1 and L2 (at the bottom). In that case, the match is considered as reliable. L3 and L4 being close to 3 different markers, this case is ambiguous: P3 and P4 are not matched with another marker.

Video / CT-scan marker matching

A set of 3D markers being reconstructed in the camera reference frame, we need to match them with the corresponding 3D points extracted from the CT-scan. We decided to use a prediction/verification algorithm [13] with additional distance constraints. This algorithm provides us not only a matching estimation, but a transformation initialization as well. The principle is to choose a triplet from the reconstructed marker set and to find a triplet from the CT-scan markers that has the same dimension (the triangle sides have approximately the same length). If no CT-scan triplet is found, a new triplet from the reconstructed marker set is chosen. When a CT-scan triplet is found to have the same dimension, we compute the rigid transformation that best matches the 2 triplets [14]. To verify this prediction, we apply this transformation to the extracted CT-scan markers and project them in the video images. Since the CT (resp. video) marker extraction algorithm provided 5% (resp. 5%) of false negatives, we accept the transformation as correct if more than 85% of the projected points are close to the extracted 2D video points. Otherwise, the 2 triplets are not homologous despite the dimension similarity (this case is illustrated on Figure 6): we have to choose a new triplet in the CT-scan marker set. In fact, to limit computation time, it is much more interesting to sort the triplet so that we first choose those whose number of potential corresponding triplets is small (typically, we begin with the triplets of largest dimension).

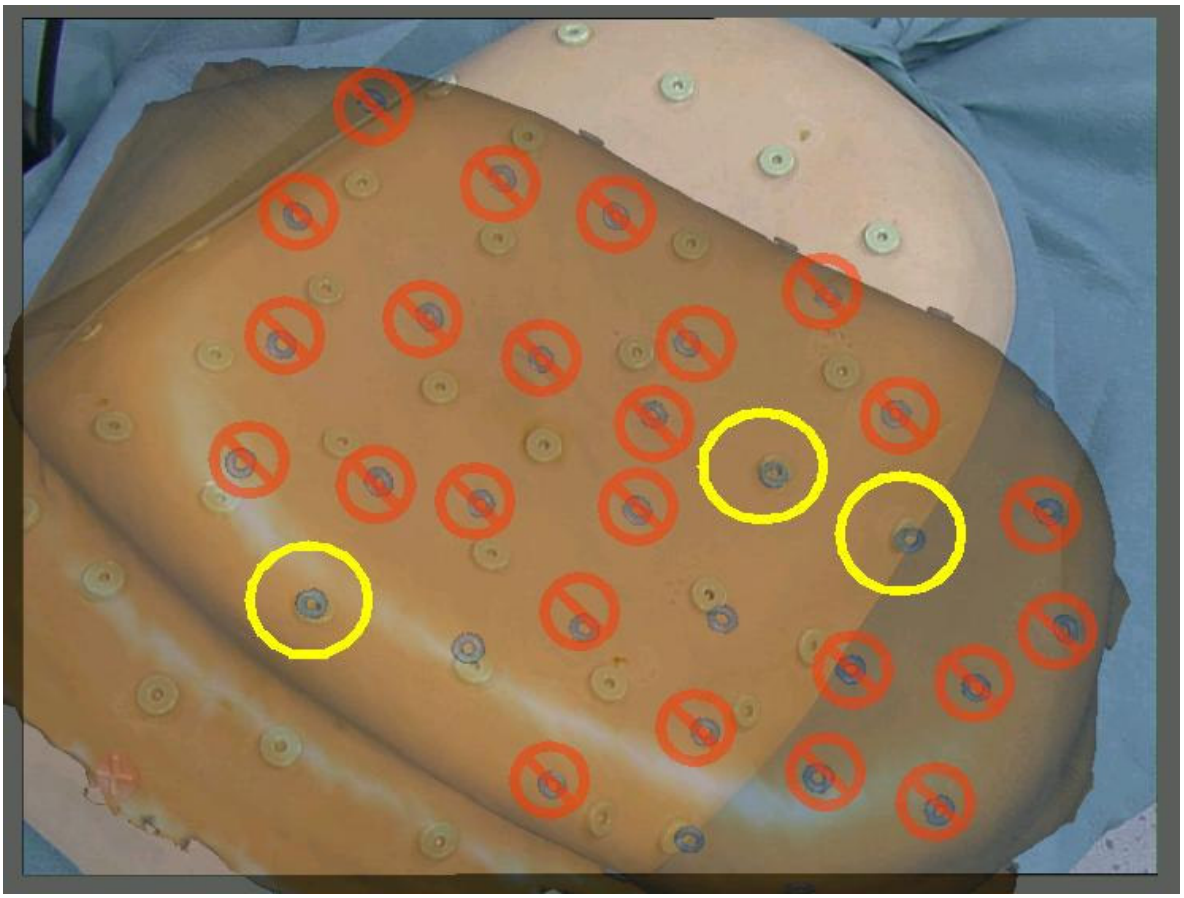

Figure 6: Illustration of a bad triplet matching. One can see in the light circles three 3D markers from the CT model that are well superimposed with three 2D markers of the video image. Since almost all the other 3D model markers are not close to video markers (in dark circles), it means that the two triplets are not representing the same markers.

When a correct transformation is found, we keep all the 3D/2D correspondences provided by the verification step to compute a more accurate transformation (see next section). In this process, markers that belong to ambiguous 2D/2D video matches are implicitly reconsidered thanks to the verification. The matching experiments made with the 210 image pairs of the phantom used in the previous section always provide a correct transformation, and 100% of the possible matches were found. The whole matching process took on average less than 2 seconds and 3 seconds at the most. Eventually, these results show that our procedure of extraction and matching is sufficiently robust and reliable for clinical use.

Computation of the best rigid transformation

We could have directly used the transformation obtained between the 2 matched triplets, but it only relies on three 3D matches. Since we need the best possible accuracy, it is desirable to take the information brought by all 3D/2D matches into account. The classical least square 3D/2D criterion supposes that the noise corrupting the 3D data is negligible. This assumption being false in our case, we optimize the maximum likelihood EPPC criterion that we developed in [15] and that is optimal in the case of noise on *both* 2D and 3D markers extraction:

$$\sum_{l=1}^{2} \sum_{i=1}^{N} \xi_i^l \cdot \frac{\left\| P^l (T * M_i) - \tilde{m}_i^l \right\|^2}{2\sigma_{2D}^2} + \sum_{i=1}^{N} \frac{\left\| M_i - \tilde{M}_i \right\|^2}{2\sigma_{3D}^2}$$

where **T** is the sought transformation, $\mathbf{P^l}$ is the projective function of the 1st camera, $\tilde{M}_i$ are the extracted 3D CT-scan markers, $\tilde{m}_i^l$ the extracted video markers from the camera **l**, σ_{2D} and σ_{3D} the estimated standard deviation of the noise corrupting the CT and video markers extraction, M_i the **exact** location of the CT markers and ξ_i^l a binary variable equal to 1 if M_i is observed by the 1st camera and 0 otherwise. Note that M_i are not known and are consequently estimated just as the transformation **T** (M_i are called hidden variable). We perform an alternated minimization w.r.t. the two groups of variables, starting from a transformation 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 becomes negligible. An experimental study realized with the phantom [15] showed that a superimposition accuracy of 2 mm was reached for targets within the liver.

Superimposition of the virtual needle on the real one

As the needle is no longer visible as soon as the radiologist introduces it under the skin, its virtual model needs to be superimposed in real-time onto the video images. This means that we have to track its location and orientation in space in the camera reference frame. To realize it, we attach an oriented square marker whose corners are automatically localized on video images in real-time using ARToolkit library [16]. Then, knowing the size of the square, we are able to localize it in the camera reference frame by minimizing the classical 3D/2D SPPC criterion [15]:

where, $\tilde{M}_i$ are the position of the 4 corners in the square marker frame, $\tilde{m}_i^l$ are the pixel position of the 4 corners in the 1^{st} video image and **T** is the transformation between the square marker frame and the camera reference frame. Calibrating the relative needle position with respect to the square marker with the pivot method [17], we are finally able to superimpose the virtual model on the real one on video images (cf. Figure 7).

$$\sum_{i=1}^{4} \left\| P^1 (T^* \tilde{M}_i) - \tilde{m}_i^1 \right\|^2 + \sum_{i=1}^{4} \left\| P^2 (T^* \tilde{M}_i) - \tilde{m}_i^2 \right\|^2$$

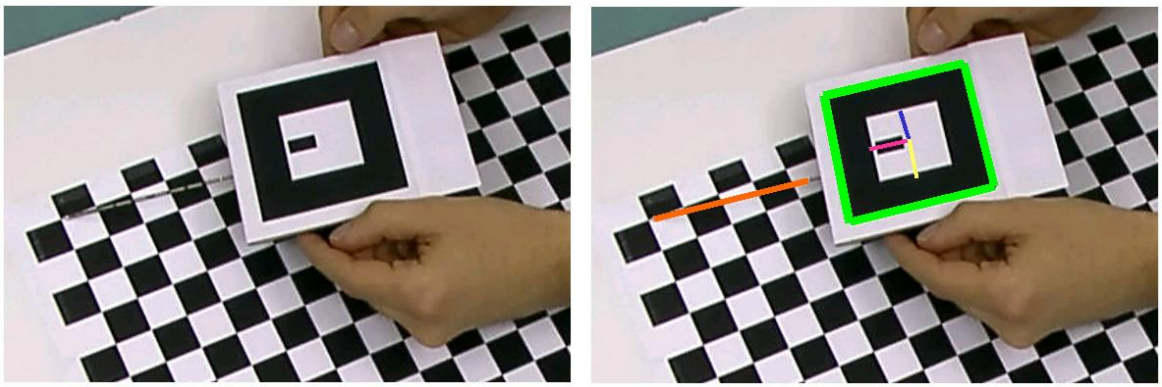

Figure 7: Superimposition of the virtual needle on the real one. Left: original image. Right: detection of the square marker position and orientation. The virtual needle is superimposed in orange.

Evaluation of the overall system

The purpose of these experiments was to assess the accuracy of the needle targeting by a surgeon using our AR guidance system. Targets were modeled with radio-opaque markers stuck on the fake liver inside the phantom. To reach one target, the augmented view provides the virtual position of the target and the needle in the two video images (cf. Figure 8). The needle orientation is guided by the color target that changes when the needle points toward the right direction. Moreover, the distance (in mm) between the virtual needle tip and the center of the virtual tumor is provided by the software. This distance varied continuously, because of the fine tremor of the hand of the operator. Two participants (a computer scientist (CS) and a surgeon) each performed 50 consecutive needle targetings of the model tumor using the augmented view. During the targeting, the operator placed the needle and stopped his movement when he thought that he had reached the center of the tumor. After each trial, the time required to position the needle was recorded, and the accuracy of the hit was verified by an independent observer using an endoscopic camera introduced into the phantom and focusing on the targets (cf. Figure 9). Results shown in Table 1 indicate clearly that our system provides enough accuracy and that it allows to reach the target very quickly with respect to the usual time needed for a standard percutaneous intervention.

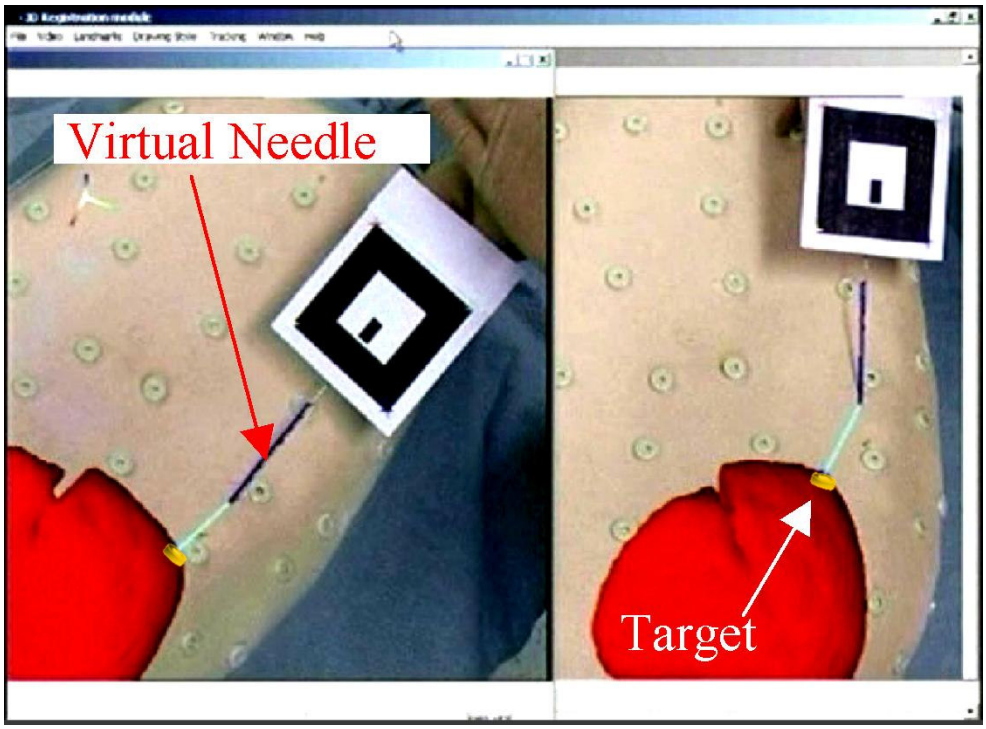

Figure 8: Augmented view seen by the surgeon on the screen. The video images display the virtual model of the target (yellow circle) and of the needle (in blue). The light blue line corresponds to the distance between the needle tip and the target.

Figure 9: Control view: an endoscopic camera is inserted inside the phantom to check the targeting accuracy.

	Total	CS	Surgeon
Average real distance (mm)	2.79 ± 1.41	3.6 ± 1.03	1.98 ± 1.27
Average time (sec)	46.57 ± 24.64	38.5 ± 21.78	54.64 ± 24.88

Table 1: Accuracy of needle targeting (mean value ± standard deviation).

Conclusion

In this paper, we have presented an AR system, based on radio-opaque marker registration, to guide the needle positioning during RF hepatic tumor ablation. To fulfill time constraints, we have developed procedures to extract and match the markers in the video and the CT-scan images. Experimental studies proved that these algorithms are robust and fast enough for clinical application. In order to evaluate the efficiency of the whole system, we conducted needle positioning experiments on targets embedded into our human abdominal phantom. Results show that our system of AR guidance is highly accurate (2.8 mm on average) and enables to locate hidden targets quickly (less than one minute on average), whereas a standard percutaneous intervention generally takes more than 5 min. The system is also easy to use, and yields reproducible results: an untrained surgeon, not involved in the initial development of the system, was able to use it repetitively with a better accuracy than the computer scientist. One other advantage of this system is the simple material needed. Indeed, only a PC, a video acquisition card, two cameras and a printed square are required, contrary to the other existing systems that need devoted tracking materials [3][6][10] (such as Optotrak, Flashpoint…) or a laser scanner [4].

In order to reach a clinical validation and use, we plan experiments on human patients to verify that our system fits real operative constraints (low organ mobility and good illumination conditions).

Bibliography

[1] McGahan JP, Dodd GD. Radiofrequency ablation of the liver: current status. American Journal of Roentgnology 2001;176(1):3—16.

[2] Gering D, Nabavi A, Kikinis R, Grimson W, Hata N, Everett P, Jolesz F, Wells W. An integrated visualization system for surgical planning and guidance using image fusion and interventional imaging. In Proc. of Second International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'99). Pages 808-819. 1999.

[3] Maurer C, Fitzpatrick J, Wang M, Galloway R, Macinuas R, Allen G. Registration of Head Volume Images using Implantable Fiducial Markers. In IEEE Transactions on Medical Imaging 1997;16:447-462.

[4] Grimson W, Ettinger G, White S, Lozano-Perez T, Wells W, Kikinis R. An Automatic Registration Method for Frameless Stereotaxy, Image-Guided Surgery and Enhanced Reality Visualization. In IEEE Transactions on Medical Imaging 1996;15(2):129-140.

[5] Peuchot B, Tanguy A, Eude M. Virtual Reality as an Operative Tool During Scoliosis Surgery. In Proc. of First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine (CVRMed'95). Pages 550—554. 1995.

[6] Merloz P, Lavallee F, Tonneti J, Pittet L. Image-guided spinal surgery: technology, operative technique and clinical practice. In *Operative Techniques in Orthopaedic.*

2000;10(1):56—63.
[7] Mitschke M, Bani-Hashemi A, Navab N. Interventions under Video-Augmented X-ray Guidance: Application to Needle Placement. In Proc. of Third International Conference on Medical Image Computing And Computer-Assisted Intervention (MICCAI'00). Pages 858--868. 2000.
[8] Guéziec A, Kazanzides P, Williamson B. and Taylor R. Anatomy-Based Registration of CT-scan and Intraoperative X-ray Images for Guiding a Surgical Robot. In *IEEE Transactions on Medical Imaging*, 17(5), October 1998.
[9] Mourgues F. Guidage par réalité augmentée: application à la chirurgie cardiaque robotisée. Phd Thesis. 2003.
[10] Langø T, Ystgaard B, Tangen G A, Hernes T A N, Mårvik R. Feasibility of 3D navigation in laparoscopic surgery. Oral presentation at the SMIT (Society for Medical Innovation and Technology) Conference, Oslo, Norway, September 5-7, 2002.
[11] Soler L, Delingette H, Malandain G, Montagnat J, Ayache N, Koehl C, et al. Fully automatic anatomical, pathological, and functional segmentation from CT scans for hepatic surgery. Computer Aided Surgery 2001;6(3):131—142.
[12] Zhang Z. A flexible new technique for camera calibration. Technical report; Microsoft Research; 1998 December
[13] Ayache N. and Faugeras O. HYPER: A new approach for the recognition and Positioning of Two-Dimensional Objects. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Volume 8 (1):44-54. 1986.
[14] Arun K. S, Huang T. S, Blostein S. D. Least-Squares Fitting of Two 3D Point Sets. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Volume 9(5):698-700. 1987.
[15] Nicolau S, Pennec X, Soler L, Ayache N. Evaluation of a new 3D/2D registration criterion for liver radio-frequencies guided by augmented reality. In IS4TM LNCS 2673; 2003 France;p.274—287.
[16] Human Interface Technology Library:
www.hitl.washington.edu/research/shared_space/download/
[17] Lavallée S, Cinquin P. and Troccaz J. Computer Integrated Surgery and Therapy: State of the Art. In C. Roux and J.L. Coatrieux, editors, *Contemporary Perspectives in Three-Dimensional Biomedical Imaging*, chapter 10, pages 239-310, IS Press, Amsterdam, NL, 1997.