Marker-Less AR in the Hybrid Room Using Equipment Detection for Camera Relocalization

Nicolas Loy Rodas, Fernando Barrera, and Nicolas Padoy

ICube, University of Strasbourg, CNRS, IHU Strasbourg, France
{nloyrodas,barreracampo,npadoy}@unistra.fr

Abstract. Augmented reality (AR) permits clinicians to visualize directly in their field of view key information related to the performance of a surgery. To track the user's viewpoint, current systems often use markers or register a reconstructed mesh with an a priori model of the scene. This only allows for a limited set of viewpoints and positions near the patient. Indeed, markers can be intrusive and interfere with the procedure. Furthermore, changes in the positions of equipment or clinicians can invalidate a priori models. Instead, we propose a marker-free mobile AR system based on a KinectFusion-like approach for camera tracking and equipment detection for camera relocalization. Our approach relies on the use of multiple RGBD cameras: one camera is rigidly attached to a hand-held screen where the AR visualization is displayed, while two others are rigidly fixed to the ceiling. The inclusion of two static cameras enables us to dynamically recompute the 3D model of the room, as required for relocalization when changes occur in the scene. Fast relocalization can be performed by looking at an equipment that is not required to remain static. This is particularly of advantage during hybrid surgeries, where an obvious choice for such an equipment is the intraoperative imaging device, which is large, can be seen in all views, but can also move. We propose to detect the equipment using a template based approach and further make use of the static cameras to speed-up the detection in the moving view by dynamically adapting the subset of tested templates according to the actual room layout. The approach is illustrated in a hybrid room through a radiation monitoring application where a virtual representation of the radiation cone beam, main X-ray scattering direction and dose distribution deposited on the surface of the patient are displayed on the hand-held screen.

Keywords: Augmented reality, RGBD cameras, Equipment detection, Camera relocalization, Radiation safety monitoring.

1 Introduction

This paper presents the development of a marker-free approach to perform mobile augmented reality (AR) in the operating theatre during hybrid surgery. By mobile AR, we refer to the display of information directly in the user's view by using a mobile screen or a head-mounted display (HMD). The targeted application is the display of 3D information related to radiation safety during X-ray

N. Navab et al. (Eds.): MICCAI 2015, Part I, LNCS 9349, pp. 463–470, 2015.
DOI: 10.1007/978-3-319-24553-9_57

guided procedures or during training sessions for such procedures. Consequently, the approach should allow a user, such as the clinician or the radio-manipulator, who can be located at different distances and positions around the patient table, to look anywhere around the table and benefit from the AR overlay. This application is subject to several constraints, such as the presence of motion in the scene, for instance the rotation and translation of the X-ray device, and the impracticability of having multiple markers visible from all interesting viewpoints. It is however important to note that precision requirements are less stringent here than in traditional image guided applications designed to guide precise gestures, for instance for tumor resection or screw placement.

The last decades have seen the development of several medical AR visualization solutions that overlay pre-operative or intra-operative patient images directly in the view of the clinician. Due to limited space, we refer the reader to [10,4] for a detailed review of these approaches and also for a description of other available visualization methods. These systems rely either on markers to locate the viewing device or on registration between a reconstructed surface and a model of the patient. For instance, the RAMP system [9] uses an inside-out approach, where a pattern of markers located near the patient is tracked with a camera placed on the HMD. In [6], an iterative closest point (ICP) matching approach is applied to register a pre-operative patient model to the view of a time-of-flight camera fixed on a mobile display. Such approaches are designed to display information related to the anatomy of the patient. In contrast, we are interested in displaying information concerning the surroundings of the patient and of the operating table, such as the skin dose and the scattered radiation map. A modeling of the scene beyond the patient is therefore required.

Several studies on radiation safety during X-ray guided procedures [11] have pointed out that clinicians and patients can be overexposed to radiation due to improper positioning of the X-ray device. Small changes in the angulation of the device can multiply the dose received by the patient or clinician by several factors. Similarly, overexposure can occur if the clinician or the lead protection devices are incorrectly positioned. In such situations, appropriate feedback could strongly reduce the received doses. To increase the awareness of clinicians w.r.t. radiation safety, [12] proposes a software that shows the radiation scattering in a virtual OR during training sessions. [3,5] propose to increase awareness by showing the dose received by the clinicians using radiation simulation and 3D person tracking. The risk is shown on a screen that overlays the radiation map over a 3D model of the room. This paper proposes a new approach to provide feedback to clinicians and staff when the X-ray device is used. It relies on mobile AR to display directly in the view of the staff key information about an upcoming X-ray image, such as the radiation beam, the main scattering direction, the radiation exposure of the patient and the current device parameters.

The AR approach that we propose uses one RGBD camera fixed on a mobile display and two static RGBD cameras fixed on the ceiling. Tracking of the moving camera is performed using an approach similar to KinectFusion [7], referred to as KinFu in the rest of this paper. One inconvenience of KinFu is the loss of tracking when motions occur in the scene or when the view changes abruptly. The static cameras are therefore a key element in the system to enable fast and convenient relocalization by keeping an up-to-date picture of the scene. We perform relocalization by detecting the same equipment simultaneously in the moving view and in at least one static view. Motivated by our clinical application, we detect the C-arm, which is large and can be partially seen in all views. We use an extended template-based detection approach based on [2]. It detects equipment *parts* for robustness and integrates natural constraints on the mobile camera location by dynamically adapting the set of tested templates according to the equipment configuration detected in the static views. This approach allows for fast relocalization in the presence of changes in the scene, as opposed to approaches like [1,8], which use a single camera and are designed for static scenes.

The contributions of this paper are as follows: we propose a KinFu based approach for mobile AR in the hybrid room using multiple RGBD cameras and equipment detection for relocalization. We also introduce a dynamic template-based approach for detection in the mobile view to fully integrate the information provided by the static cameras. To the best of the authors' knowledge, this is the first time that equipment detection is attempted in the OR using vision. Finally, we suggest a new application for mobile AR in the OR, namely the monitoring of patient dose and main direction of scatter during X-ray guided procedures.

2 Methods

The setup is composed of three synchronized and calibrated RGBD cameras (Asus Xtion Pro): two are rigidly mounted to the ceiling of the operating room and a third one is attached to the tracked hand-held display.

2.1 System Setup

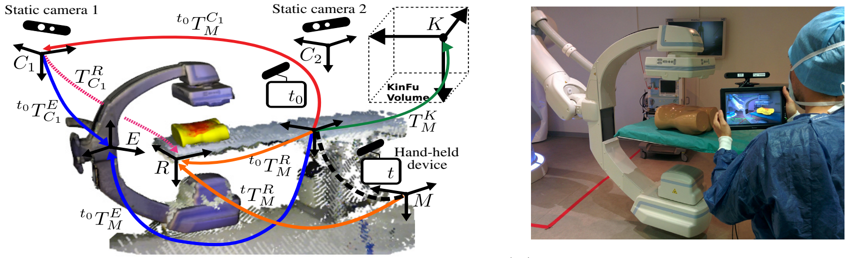
Fig. 1 presents the proposed system setup and the different transformations involved. In the calibration procedure, the transformations $T_{C_i}^R$ between each static camera ($\mathbf{C}_i$) and the room ($\mathbf{R}$) are obtained using the procedure presented in [5]; these remain constant for each setup.

The mobile device ($\mathbf{M}$) is tracked using KinFu, yielding at each time step t a relative pose ${}^{t}T_M^K$ with respect to the virtual KinFu volume ($\mathbf{K}$) generated during tracking. The simultaneous detections at t_0 of a piece of equipment $\mathbf{E}$ in at least one static camera ($\mathbf{C_1}$ in Fig. 1a) and in the moving view, provide object-to-camera transformations used for computing an initial registration ${}^{t_0}T_M^R$ of the hand-held device with respect to the room coordinate system $\mathbf{R}$ in the following manner:

$$ {}^{t_0}T_M^R = {}^{t_0}T_M^E \, ({}^{t_0}T_{C_1}^E)^{-1} \, T_{C_1}^R. \quad (1) $$

During tracking, ${}^{t_0}T_M^R$ is applied to the current relative transformation ${}^{t}T_M^K$ to obtain at each time step the pose of the moving device with respect to the room:

$$ {}^{t}T_M^R = {}^{t}T_M^K \, (T_M^K)^{-1} \, {}^{t_0}T_M^R, \quad (2) $$



(a) Setup and corresponding transformations.

(b) Hand-held device displaying the AR visualization.

Fig. 1. System setup.

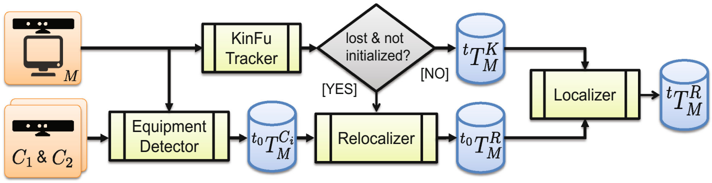


Fig. 2. Tracking pipeline.

where T_M^K is a constant transformation in the system. When tracking is lost, camera relocalization is achieved by recomputing ${}^{t_0}T_M^R$ using new simultaneous equipment detections in both views.

2.2 Tracking Pipeline

An overview of the tracking pipeline is shown in Fig. 2. All processes are executed concurrently in different threads.

KinFu Tracker: This process makes use of an open-source implementation of KinectFusion [7] to track the motion of **M**. It uses the GPU to simultaneously construct a 3D map of the environment and to estimate with ICP the transformation relating **M** to **K**. At each time step, the ICP error is compared to a threshold in order to detect any tracking failure.

Equipment Detector: This process performs equipment detection in each view and is further described in Sec. 2.3. Detecting an object in the moving view alone is not sufficient for camera relocalization due to the possible ambiguities caused by the fact that the object can also be moving.

Relocalizer: This process corresponds to the equipment-based relocalization algorithm. It uses simultaneous equipment detections to compute ${}^{t_0}T_M^R$ at initialization or to recompute it when **M** is lost. The transformation from Eq. 1 is refined by applying ICP between the point cloud of the moving camera and the merged point cloud from the static cameras, restricted to a large area around the object.

Localizer: This process applies ${}^{t}T_M^K$ and ${}^{t_0}T_M^R$ to compute the pose ${}^{t}T_M^R$ of the mobile camera w.r.t. the room using Eq. 2 for the AR visualization.

2.3 Equipment Detection

We apply a multimodal template matching approach inspired by [2] for detecting equipment in each of the views and estimating the cameras' relative poses w.r.t. the detected equipment. The templates sample the possible appearances of an object and are built from densely sampled color gradients and depth map normals. Because each template is labeled with the corresponding relative camera-object transformation, when a template is found at test time, a coarse estimation of the object's pose is also provided. This pose is further refined using ICP.

Template Generation: Since CAD models of the equipment from an operating room are not easily available, we first generate a 3D model by scanning the object with a RGBD camera. Then, as in [2], we equally sample a set of viewpoints with respect to the coordinate center of the model by recursively dividing the space into a polyhedron. Each obtained vertex is considered as a camera viewpoint from where virtual color and depth images are generated and used for computing the templates. This is repeated for polyhedrons with different sizes to generate images at different scales. In [2], the assumption is made that the objects to detect are placed over a planar surface. Therefore, viewpoints are only sampled from the upper hemisphere of the model. They are also only sampled using a camera parallel to the horizon and looking at the object's center. In our case, the cameras are not necessarily horizontal and have very different viewpoints. Furthermore, the C-arm has a large set of possible orientations. To generate templates for all possible situations, virtual images are first generated from views sampled from a sphere around the object (Fig. 3a) using a horizontal camera pointing towards the object's centre. The same procedure is then repeated for a discrete set of n 3D orientations of the object $\Theta = \{\theta_1, \ldots, \theta_n\}$. Since objects are not constantly moving in an OR, the slower detection speed caused by the larger set of templates is not a limitation for detection in the static views. However, faster detection is required in the moving view for a comfortable user experience. We therefore dynamically adapt the set of templates tested in the moving camera based on the object pose detected in the static views, as described below.

Dynamic Template Subset Selection: We make use of the detected 3D object position in the static cameras to gain information about the object poses that the moving camera should expect. We include the additional assumption that the moving device is constrained to move within a certain part of the room and between pre-defined heights. The set of relevant templates is thereby drastically reduced. In practice, we precompute for each 3D orientation $\theta \in \Theta$ the subset S_θ of potentially visible templates (Fig. 3a). At test time, the static cameras detect the current object orientation and the system dynamically loads the corresponding template subset to be used by the moving camera detector. This procedure reduces the number of computations per frame and accelerates the tracking initialization and relocalization procedures.

(a) Sampling procedure. Red poses form a template subset.
(b) Detected equipment overlaid in static view.
(c) Part detected and overlaid in moving view.

Fig. 3. Equipment detection: template generation and result examples.

Part-Based Detection: A known drawback of template matching approaches is the fact that the complete template must be visible in the image for it to be detected. This can become an issue when dealing with large objects or close views of the object. We use a part-based detection approach that copes with these limitations by looking for separate parts of the same object in the images. First, we divide the object's model into a finite number of parts to obtain separate models sharing the same reference coordinate system. Then, templates are generated for each part using the procedure described above. Detection is performed separately per part for finding the complete object even when this one is partially occluded or not fully visible. Making the assumption that the parts located higher than the table are the most visible, the parts are tested by decreasing order of heights, making use of the equipment pose estimated by the static cameras.

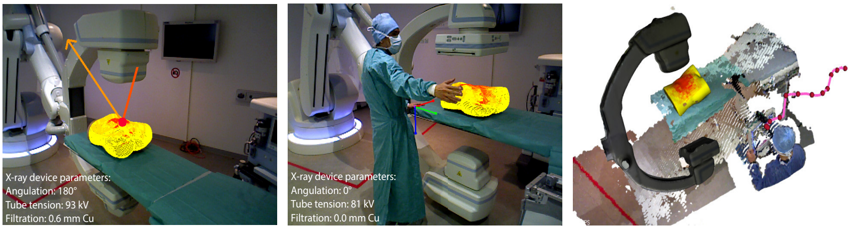
Mobile AR Visualization: The estimation of ${}^{t}T_{M}^{R}$ allows us to overlay virtual elements directly into the user's point of view. We do so by back-projecting the virtual objects directly into the RGB image, shown on the moving screen. We make use of the depth map from the moving camera to detect when the virtual object is occluded and display only its parts visible in the current view.

3 Results

To validate our system, we recorded several sequences illustrating its typical use and several challenges that can occur, such as C-arm rotation, abrupt changes of viewpoint orientation, large changes in the position of the user around the table, clinicians walking in the view and other types of occlusions. The medical imaging system used in the room is a Siemens Zeego. Note, however, that the kinematic information was not available. Some of the results are described below. They are more comprehensively presented in the supplementary video[1], which also includes live recordings, since video material is better suited for evaluation.

The approach is tested using a computer equipped with an i7-3930K 6-core processor along with a GeForce GTX Titan GPU. Parallelization occurs through

[1] Supplementary video can be found at: `http://camma.u-strasbg.fr/videos/`



(a) AR visualization showing the main direction of scatter (orange), patient exposure and current device angulation.

(b) AR visualization in the moving view where depth is used for consistent display in case of occlusion.

(c) Hand-held device trajectory displayed in the fused point cloud obtained from the static cameras.

Fig. 4. Illustration of the AR and tracking results for our clinical application.

multi-threading as described in Sec. 2.2 and through the use of the GPU for KinFu. In these experiments, the radiation maps are pre-computed offline using the approach presented in [5]. The patient is represented by an anatomical phantom, whose 3D shape has been scanned. Its exposure to radiation is displayed by texturing the phantom's surface according to the dose value at each 3D location for the current device configuration. Radiation scattering is caused by the interactions of the radiation with the surface of the patient. The main direction of scatter is computed from the position and orientation of the X-ray tube provided by the equipment detection.

Fig. 3 shows examples of object detections in the moving and static views. The 3D models are overlaid over the original images. In Fig. 3c, only the top part of the C-arm was detected. With the parameters used in our experiments, the dynamic template selection permits to divide the number of tested templates in the moving view by a factor of 10. Two examples of visualizations seen by the user are shown in Fig. 4 for different layouts of the room and positions. They illustrate the robustness of the system to occlusions and also the benefits of using the depth for consistent display of the virtual information in such situations. The overall system, including the visualization, runs at a framerate of 27 fps. The tracked path of the moving camera is shown in Fig. 4c inside the merged point clouds from the static cameras.

4 Conclusion

We have presented a mobile augmented reality approach for radiation awareness in the hybrid room. It relies on KinFu for camera tracking and on equipment detection for relocalization. By using three RGBD cameras, it has the advantage of being flexible and non-intrusive. The system is a proof-of-concept demonstrated on an important medical use-case. It is currently limited by the fact that the computation of the radiation risk map is not real-time and that the patient is not tracked. Future work will need to address these limitations. This system is however useful in its current form for radiation awareness training and can also be employed for other AR applications.

Acknowledgements. This work was supported by French state funds managed by the ANR within the Investissements d'Avenir program under references ANR-11-LABX-0004 (Labex CAMI), ANR-10-IDEX-0002-02 (IdEx Unistra) and ANR-10-IAHU-02 (IHU Strasbourg).

References

1. Glocker, B., Izadi, S., Shotton, J., Criminisi, A.: Real-time rgb-d camera relocalization. In: ISMAR, pp. 173–179 (2013)
2. Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N.: Model based training, detection and pose estimation of texture-less 3D objects in heavily cluttered scenes. In: Lee, K.M., Matsushita, Y., Rehg, J.M., Hu, Z. (eds.) ACCV 2012, Part I. LNCS, vol. 7724, pp. 548–562. Springer, Heidelberg (2013)
3. Ladikos, A., Cagniart, C., Ghotbi, R., Reiser, M., Navab, N.: Estimating radiation exposure in interventional environments. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 237–244. Springer, Heidelberg (2010)
4. Linte, C., Davenport, K., Cleary, K., Peters, C., Vosburgh, K., Navab, N., Edwards, P., Jannin, P., Peters, T., Holmes, D., Robb, R.: On mixed reality environments for minimally invasive therapy guidance: Systems architecture, successes and challenges in their implementation from laboratory to clinic. Comput. Med. Imaging Graph. 37(2), 83–97 (2013)
5. Loy Rodas, N., Padoy, N.: Seeing is believing: increasing intraoperative awareness to scattered radiation in interventional procedures by combining augmented reality, monte carlo simulations and wireless dosimeters. International Journal of Computer Assisted Radiology and Surgery (2015)
6. Maier-Hein, L., Franz, A., Fangerau, M., Schmidt, M., Seitel, A., Mersmann, S., Kilgus, T., Groch, A., Yung, K., dos Santos, T., Meinzer, H.P.: Towards mobile augmented reality for on-patient visualization of medical images. In: Bildverarbeitung für die Medizin, pp. 389–393. Springer (2011)
7. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.W.: Kinectfusion: Real-time dense surface mapping and tracking. In: ISMAR, pp. 127–136 (2011)
8. Salas-Moreno, R.F., Newcombe, R.A., Strasdat, H., Kelly, P.H.J., Davison, A.J.: SLAM++: simultaneous localisation and mapping at the level of objects. In: CVPR, pp. 1352–1359 (2013)
9. Sauer, F., Khamene, A., Bascle, B., Rubino, G.J.: A head-mounted display system for augmented reality image guidance: Towards clinical evaluation for imri-guided neurosurgery. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 707–716. Springer, Heidelberg (2001)
10. Sielhorst, T., Feuerstein, M., Navab, N.: Advanced medical displays: A literature review of augmented reality. Journal of Display Technology 4(4), 451–467 (2008)
11. Vanhavere, F., Carinou, E., Gualdrini, G., Clairand, I., Merce, M., Ginjaume, M.: The oramed project: Optimisation of radiation protection for medical staff. IFMBE Proceedings, vol. 25/3, pp. 470–473 (2009)
12. Wagner, M., Dresing, K., Wolfram, L., Ahrens, C.A., Bott, O.J.: Siscar-gpu: fast simulation and visualization of intraoperative scattered radiation to support radiation protection training. MIE 180, 968–972 (2012)