Cost-effective Techniques for Patient Positioning in Percutaneous Radiotherapy Using Optical Imaging Systems

Hamid Sarmadi
Artificial Vision Applications Group
IMIBIC
Córdoba, Spain
hamid.sarmadi@imibic.org

***Abstract*—Patient positioning is an important task in radiation therapy the prevents healthy tissue to be radiate with dangerous electromagnetic waves. There are already different systems that assist precise patient positioning however they expose the patient to extra radiations, are not accurate enough, or are to expensive for the average hospital in Spain. In this thesis we want to verify the possibility of creating a system based on recently developed computer vision algorithms and RGBD sensors that is affordable and at the same time accurate for the purpose of patient positioning.**

Keywords—radiotherapy; radiation therapy; computer vision; depth sensor; rgbd sensor; patient positioning; pose estimation

I. Thesis Information

Contact Information:
hamid.sarmadi@imibic.org

Thesis Directors:
Rafael Muñoz-Salinas, Rafael Medina-Carnicer

University Department:
Departamento de Informática y Analisis Numérico
Universidad de Córdoba
Córdoba, Spain

Starting Date:
April 2017

II. Summary and Relevance

Percutaneous radiotherapy is done by focusing an electromagnetic wave at the tumor or the target tissue in the body. It has two main phases. First, the planning phase in which normally a CT scan, PET scan or MRI are done to determine the exact position of the target tissue in the patient's body and acquiring a 3D model. Second, the treatment phase, where there are normally multiple session and in each one the patients needs to be positioned in the exact same pose in which the initial scans are done. Then with the help of the initial scans the exact position of the target area in the body is determined and is treated with an electromagnetic beam.

There could always be errors in positioning patient in the exact pose as in the initial scan and that there also could be movements in the internal organs. These can cause the radiation of healthy tissue in radiation therapy which begs for accurate validation systems for patient positioning. Furthermore, recently, new methods such as intensity modulated radiation therapy (IMRT) that can achieve high precision and volumetric modulated arc therapy (VMAT) which can reduce the time of radiation have been created that put extra reliance on precises positioning of the patient.

There are already high precision patient positioning methods. One example works using X-ray imaging to determine the bone structure at the time of treatment in two dimensional. Then the digitally reconstructed radiograph (DRR) from CT scans obtained in the planning phase are manually aligned with the 2D X-ray image to estimate the position of the tumor [6]. This method however puts the patient under extra radiation for each session of the therapy and is not very accurate. Another X-ray based approach is computed tomography cone beam (CBCT). This method is more accurate since it creates a 3D reconstruction of the bone structure by rotating a cone beam around the patient and combining its images. The 3D bone reconstruction then is compared to the CT scan. Though more accurate, the method still uses X-ray which puts healthy tissue under dangerous radiation. Further more it has a high cost.

Other radiation-less methods have been introduced. In [7] a method based on orthogonal video sequences is proposed. The videos are aligned with reference images to estimate the pose. This technique has been used for the treatment of head and neck area [8]. In another type of approach reflective markers in conjunction with infrared cameras have been utilized [9]. Reflective markers are placed in pre-know positions on patient's body and multiple infrared cameras are used to determine the 3D position of each marker and hence estimate the pose of the patient. However, this technique can only estimate the pose of the markers and not the whole surface of the body. In more recent years, 3D reconstruction has gained more popularity. One method is to project a beam of laser on the patient's body and using the images taken with calibrated cameras reconstruct the 3D surface of the body. This has been applied for rectal cancer [10] prostate cancer [11] patients. By project 2D random patterns instead of a one dimensional beam [12] it is possible to have a real time 3D reconstruction of patient's position. This system can also be use to give the patient feedback regarding the accuracy of their position during

This thesis is funded under the project IFI16/00033 (ISCIII) from Spain.

the radiation therapy [1]. This type of method could obtain a high precision [2].

More specifically CATALYST [1,2] and ALIGNRT [3] are two industrial level approaches where the visual 3D scan of the patient is matched with the CT scan done in the planning phase for verification of the patient's position. These methods, however, have a high cost for the average hospital in Spain.

On the other hand, in the field of computer vision, improvements have been done in human pose estimation using a single view (monocular) or multiple cameras. M. J. Marín-Jiménez from our group has published multiple methods for monocular human pose estimation in collaboration with Oxford University and ETH Zurich [13-15]. These methods are based on appearance based detectors and pictorial structures [16] probabilistic model of human poses. There are also multi-view approaches. One variance of them is to extend single view methods to multiple cameras and use the extra 3D information to improve the pose estimation. There have been publications from our group also in this area [17,18]. In [5] the pictorial structures is extended to three dimensions and random forest based detectors have been used as fast appearance models. They can obtain good results in pose estimation of football players. Marker-less motion capture algorithms are another class of methods that work based on a pre-known model of the subject to be tracked [19,20]. Our group has also experience in the field of marker-less motion capture [21,22]. Most recently over the counter depth cameras have been used for the purpose of 3D human pose estimation [23]. These are similar to the already mentioned CATALYST and ALIGNRT systems with the difference that they utilize projected patterns with visible light whereas normally depth cameras such as Microsoft Kinect use the infrared light hence they would not pose limitations on environment lighting for example in the radiation therapy treatment room.

In this thesis, we intend to verify the possibility of obtaining an affordable visual based patient positioning system that combines advanced computer vision approaches with consumer level depth sensors and cameras with acceptable precision. Our hypothesis is that we can use recent advances in computer vision algorithms in the fields of 3D reconstruction and human pose estimation combined with recently commercialized affordable sensors to create a patient positioning system for radiation therapy. The current technologies do not use the latest algorithms and are also expensive. We suspect that it is possible to create an affordable patient positioning system with our approach to the problem. More precisely our goals in this thesis are:

1- Designing a portable computer vision based portable patient positioning system for radiation therapy.

2- Incorporating 3D reconstruction algorithms in conjunction with the new affordable consumer level 3D sensors in the system.

3- Employing human pose estimation and/or marker-less motion capture algorithms to help determine the precise position of the target.

A. *Designing a portable computer vision based portable patient positioning system*

Radiation therapy room normally have ceilings where it is possible to install specialized equipment and wiring. In systems such as CATALYST and ALIGNRT the 3D scanners need to be installed in special positions on the ceiling so that they do not interfere with other equipment. Even after installation they need to be calibrated in a slow process by a specialized person.

We intend to design a system that could be easily installed and detached from the room and could be calibrated quickly and easily. This will facilitate moving the system from one hospital to another hospital, and reduce the installation and maintenance costs.

We would use affordable consumer level 3D cameras similar to Microsoft Kinect that in addition to being a 3D scanner provide us with high resolution color images which could help significantly in the task of 3D reconstruction and pose estimation compared to using only depth sensors. Furthermore such cameras make it possible to to calibrate the system in a short amount of time. In our group, a fast calibration algorithm using fiducial planar markers has been already developed [4] and is included in the well-known OpenCV computer vision library. We would use the same type of fiducial markers and attache them to the stretcher where the patient is positioned, visible to the cameras for the purpose of fast calibration.

B. *Incorporating 3D reconstruction algorithms in conjunction with the new affordable consumer level 3D sensors*

In the past few years after the release of the initial Kinect sensor, which is the combination of a color digital camera and a depth sensor i.e. RGB-D, new computer vision algorithms have been developed for 3D reconstruction that take advantage of the depth data or both depth and image data at the same time, notably starting with the KinectFusion algorithm. While there have been a lot of improvements in these algorithms for different types of scenes the accuracy of reconstruction is still not as good for raw usage in our patient positioning system. We would like to get help from the visual information that comes from the color camera and other techniqures to improve upon the existing algorithms and make a 3D reconstruction system that is fast but also accurate and suitable for the application we have in mind.

C. *Employing human pose estimation and/or marker-less motion capture algorithms to help determine the precise position of the target*

Pictorial Structures is one of the most successful human pose estimation algorithms that is normally used in single view scenarios. However, this method is normally useful for pose estimation of humans that are standing. Recently, there have been developments in expanding this algorithm to combine information from multi-views [5]. Our group has experience with this type of algorithm.

Additionally we intend to build upon the current human motion capture algorithms that is developed in our group. We would extend them by incorporating the extra 3D information that we get from the sensors. This technique requires a model of the human to start its tracking. Using the 3D model that could be obtained from the initial scan in the planning phase of the treatment (e.g. CT scan, MRI, etc.) can help the motion capture algorithm significantly.

Normally, marker-less motion capture systems rely on inferring 3D information from the multiple view normal cameras. We think that we can improve the algorithm by additionally taking advantage of the depth information from the RGB-D sensors, in both initialization of the pose and tracking of the pose. More specifically in the tracking phase, it would be possible to track non-rigid deformations of the body accurately with the help of the depth sensor. For example, consider the up and down movement of the chest that happens with respiration. It would be of high advantage for accuracy if we were able to track the deformable surface of the chest.

III. METHODOLOGY AND WORK PLAN

A. Designing a Portable Imaging system

The first step in doing the thesis is to design a self calibrating imaging system that could be easily ported and is sufficiently accurate at the same time. In the past few years depth sensors that take advantage of structured active lighting to estimate a distance image for the devices have been commercialized for the purpose of entertainment (e.g. Microsoft Kinect) and research. These are affordable devices that normally project an infrared random pattern on the environment an calculate the depth for the pixels of their infrared camera by triangulation. Not using visible light - unlike what is done in CATALYST and ALIGNRT – removes the need for special considerations with regard to lighting of the indoor environment. Furthermore, easy calibration procedure removes the need of specialized personel for installation and facilitates transfering the system between rooms or other hospitals.

For this purpose we have done a research for automatic pose estimation and calibration of a multi-camera system with respect to a set of ArUco markers. ArUco markers are visual planar fiducial markers that can be easily attached to flat surfaces and work as a reference in the 3D environment. This system is able to estimate the pose of the ArUco markers with respect to each other, estimate the relative pose of the cameras with respect to each other and also track the relative pose between the marker and cameras at each frame all at the same time. In this work were able to reach sub-millimeter accuracy in estimating the relative pose between the cameras (which are fixed relative to each are) and the set of markers (which are also fixed with respect to each other). This is despite the fact that we used low resolution color cameras and we did not even use depth sensors. This was a promising mile stone in this phase of thesis. You can see some of our results in Fig. 1 where we used 3 cameras in a circle configuration and the markers were attached to an object in the middle of the circle.

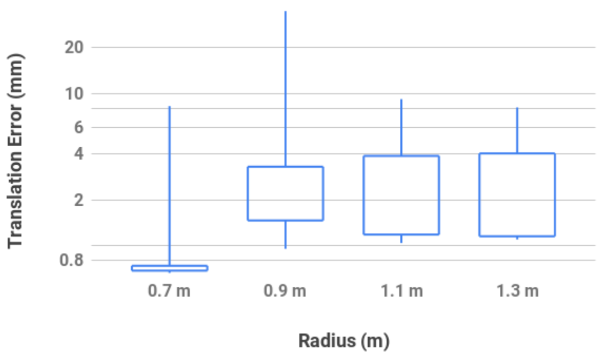

Figure 1: Error in estimation of relative position between the cameras and the markers using 3 cameras.

The next milestone in this phase would be employing the RGB-D sensors taking into account both depth and color image for estimating the 3D surface of the patient. We are intending to design an algorithm that could effectively combine the dense information that comes from the depth cameras with sparse information that comes from planar markers or feature points in the color camera for accurate tracking of the cameras and reconstruction of patient's body surface.

B. Incorporating Human Pose Estimation

The second phase of this thesis would be to use the accurate imaging system that is designed in the first phase and give its output to a human pose estimation system. Human pose estimation can determine the skeletal pose of the patient from the color image or the depth image. In order to be able to estimate the human pose accurately we need to adapt existing algorithms so that they can take our surface reconstruction as input. We are not aware of any system in patient positioning that incorporates human pose estimation.

The first milestone in this phase is a human pose estimator that could possibly take a model of the patient as input for higher accuracy pose estimation and use the surface reconstruction from the previous phase for a precise human pose estimation.

After that we need to use the information from our pose estimator and combine it with the 3D model from patient's CT scan. This is needed to model patient's inside organs and predict their position in the 3D space. Therefore, second milestone in this phase would be high accuracy boundary estimation of the area in the body that needs to be treated with radiation therapy.

REFERENCES

[1] D. J. D'Ambrosio, J. Bayouth, I. J. Chetty, M. K. Buyyounouski, R. A. Price Jr, C. R. Correa, T. J. Dilling, G. E. Franklin, P. Xia, E. E.R. Harris, A. Konski, "Continuous localization technologies for radiotherapy delivery: Report of the American Society for Radiation Oncology Emerging Technology Committee," Practical Radiation Oncology, 2:145-150, 2012

[2] P.J. Schöffel, W. Harms, G. Sroka-Perez, W. Schlegel, C.P. Karger, "Accuracy of a commercial optical 3D surface imaging system for realignment of patients for radiotherapy of the thorax," Phys Med Biol, 52:3949-3963, 2007

[3] F. Walter, P. Freislederer, C. Belka, C. Heinz, M. Söhn and F. Roeder, "Evaluation of daily patient positioning for radiotherapy with a commercial 3D surface-imaging system (Catalyst™)", Radiation Oncology, 11:154, 2016

[4] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, M. J. Marín-Jiménez "Automatic generation and detection of highly reliable fiducial markers under occlusion," Pattern Recognition, 47(6):2280-2292, 2014

[5] V. Kazemi, M. Burenius, H. Azizpour, and J. Sullivan. "Multi-view Body Part Recognition with Random Forests," BMVC 2013

[6] S. Yoshino, K Miki, K Sakata, Y Nakayama, K Shibayama, S. Mori, "Digital reconstructed radiography with multiple color image overlay for image-guided radiotherapy," Journal of Radiation Research, 56(3):588-593, 2015

[7] L.S. Johnson, B.D. Milliken, S.W. Hadley, et al., "Initial clinical experience with a video-based patient positioning system," Int J Radiat Oncol Biol Phys, 45:205–213, 1999

[8] B.D. Millikin, S.J. Rubin, R.J. Hamilton, et al., "Performance of a video-image-subtraction-based patient positioning system," Int J Radiat Oncol Biol Phys, 38:855–866, 1997

[9] G. Baroni, G. Ferrigno, R. Orecchia, et al., “Real-time three-dimensional motion analysis for patient positioning verification,” Radiother Oncol, 54:21–27, 2000

[10] C. Moore, F. Lilley, V. Sauret, et al., “Opto-electronic sensing of body surface topology changes during radiotherapy for rectal cancer,” Int J Radiat Oncol Biol Phys, 56:248–258, 2003

[11] R.I. MacKay, P.A. Graham, J.P. Logue, et al., “Patient positioning using detailed three-dimensional surface data for patients undergoing conformal radiation therapy for carcinoma of the prostate: A feasibility study” Int J Radiat Oncol Biol Phys, 49:225-230, 2001

[12] J.P. Siebert, S.J. Marshall, “Human body 3D imaging by speckle texture projection photogrammetry,” 20:218-226, 2000

[13] M. Eichner, M. J. Marín-Jiménez, A. Zisserman, V. Ferrari, “2D Articulated Human Pose Estimation and Retrieval in (Almost) Unconstrained Still Images” International Journal of Computer Vision 99(2): 190-214 (2012)

[14] V. Ferrari, M. J. Marín-Jiménez, A. Zisserman, “Progressive search space reduction for human pose estimation” CVPR 2008

[15] V. Ferrari, M. J. Marín-Jiménez, A. Zisserman, “Pose search: Retrieving people using their pose,” CVPR 2009

[16] P. Felzenszwalb, D. Huttenlocher. “Pictorial Structures for Object Recognition,” International Journal of Computer Vision 61(1): 55-79, 2005

[17] R. Muñoz-Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, A. Carmona-Poyato, “Particle filtering with multiple and heterogeneous cameras,” Pattern Recognition, 43 (7): 2390-2405, 2010

[18] R. Muñoz-Salinas, “A Bayesian plan-view map based approach for multiple-person detection and tracking,” Pattern Recognition, 41 (12), 3665-3676, 2008

[19] J. Bandouch, F. Engstler and M. Beetz, “Accurate Human Motion Capture Using an Ergonomics-Based Anthropometric Human Model,” Lecture Notes In Computer Science, 5098: 248 – 258, 2008

[20] A. Sundaresan and R. Chellappa, “Model Driven Segmentation of Articulating Humans in Laplacian Eigenspace,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 30: 1771-1785, 2008

[21] E. Yeguas-Bolivar, R. Muñoz-Salinas, R. Medina-Carnicer, A. Carmona-Poyato, “Comparing evolutionary algorithms and particle filters for Markerless Human Motion Capture,” Applied Soft Computing, 17:153-166, 2014

[22] A. Cano, E. Yeguas-Bolivar, R. Muñoz-Salinas, R. Medina-Carnicer, S. Ventura, “Parallelization strategies for markerless human motion capture,” Journal of Real-Time Image Processing, 17:1-15, 2014

[23] M. Stommel, M. Beetz, W. Xu, “Model-free detection, encoding, retrieval, and visualization of human poses from kinect data,” IEEE/ASME Transactions on Mechatronics, 20 (2): 865-875, 2015