PhD forum: Volumetric 3D Reconstruction Without Planar Ground Assumption

Hadi Aliakbarpour and Jorge Dias
Institute of Systems and Robotics, University of Coimbra.
{hadi,jorge}@isr.uc.pt

Abstract—**This paper proposes a framework to perform volumetric 3D reconstruction using a camera network. A network of cameras observes a scene and each camera is rigidly coupled with an Inertial Sensor (IS). The 3D orientation provided by IS is used firstly for definition of a virtual camera network whose axis are aligned to the earth cardinal directions. Then a set of virtual planes are defined for the sake of 3D reconstruction with no planar ground assumption and just by using 3D orientation data provided by IS. A GPU-based implementation of the proposed method is provided to demonstrate the promising results.**

I. Introduction

Recently, many efforts have been put in providing new approaches for human 3D reconstruction using a camera sensor network, specially without using any markers. The motivation of this work is to provide a multi-purpose framework which is able to register 3D information from a set of 2D images by exploiting inertial planes and without having any planar ground assumption. The use of inertial sensors to accompany computer vision applications is recently attracting attentions of the researchers. Nowadays, IS has become much cheaper and more available. Thanks to the availability of MEMS chipsets, there are many mobile-phones which are equipped with this sensor and camera as well. In our setup, we assume that each camera within the network is rigidly attached to an IS. The 3D orientation provided by the IS is used firstly to define a virtual camera instead of each IS-camera couple. As a sub-consequence the extrinsic calibration among these virtual cameras will be relaxed just to a 3D translation vector which can make the calibration process less difficult and more robust[5], [3], [7]. Secondly, the IS's 3D orientation data is used to define a set of virtual planes in the scene for the sake of 3D volumetric reconstruction. These planes are all horizontal and they are virtual in the sense that no planar ground assumption is used.

II. Data registration

The idea is to use a network of cameras to observe the scene. Each camera within the network is rigidly coupled with an IS. Using fusion of inertial and visual information it becomes possible to consider a virtual camera instead of each couple. Such a virtual camera has a horizontal image plane and its optical axis is parallel to the gravity and is downward-looking. As a result, the image plane is aligned with respect to the earth cardinal directions (North-West-Gravity). Fig. 1 shows a network of such virtual cameras. In order to obtain an image plane of a virtual camera , a homography-based approach described in [3], [4] has been used which fuses inertial data from IS and image plane of real camera to produce the corresponding virtual camera's image plane. As described in [3], [4], the homography matrix which transforms the real camera image plane to its corresponding virtual camera image plane can be expressed as following:

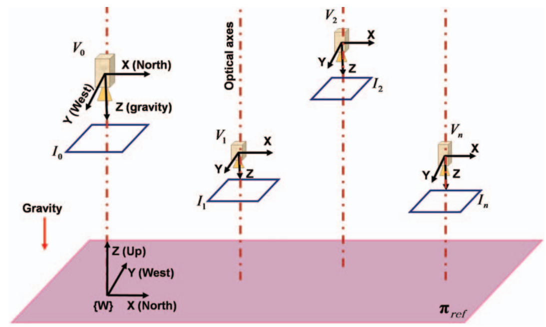

Figure 1. A network of virtual cameras: The coordinate frames of all virtual cameras are aligned to the world reference frame.

$$^{V}H_{C} = K\,^{V}R_{C}\,K^{-1} \tag{1}$$

where $^{V}R_{C}$ is the rotation matrix among the real and virtual cameras[3]. The way of obtaining $^{V}R_{C}$ is explained in [3].

By taking the advantage of inertial data, a horizontal world plane π_{ref} , which is common between all virtual cameras, has been defined in the world reference frame$\{W\}$ (see Fig. 1). As mentioned, the idea is to register virtual image data on the reference plane π_{ref} . The reference 3D plane π_{ref} is defined such a way that it spans the X and Y axis of $\{W\}$ and it has a normal parallel to the Z. In this proposed method the idea is to not using any real 3D plane inside the scene for estimating homography. Hence we assume there is no a real 3D plane available in the scene so our $\{W\}$ becomes a virtual reference frame and consequently π_{ref} is a horizontal virtual plane on the fly. The details of the world reference frame can be found in [4].

After obtaining the virtual camera's image plane it is desired to find a homography matrix $^{\pi}H_{V}$ that transforms points from the virtual image plane I' to the common world 3D plane π_{ref} (recalling that these two planes are defined to be parallel). Here we continue to formally define such a homography matrix using the rotation and translation between these two planes (I' and π_{ref}). A 3D point $\mathbf{X} = [\ X \quad Y \quad Z \quad 1\ ]^{T}$ lying on π_{ref} can be projected onto virtual image plane as

$$\mathbf{x} = \,^{\pi_{ref}}H_{v}\,\mathbf{X} \tag{2}$$

where $^{\pi_{ref}}H_{v}$ is a homography matrix which maps the π_{ref} to the virtual image plane and is defined by

$$^{\pi_{ref}}H_{v} = K[\ \mathbf{r}1 \quad \mathbf{r}2 \quad t\ ] \tag{3}$$

Hadi Ali Akbarpour is supported by the FCT (Portuguese Fundation for Science and Technology). The authors would like to thank Luis Almeida for his contribution in the smart-room experiments implementation.

978-1-4577-1707-9/11/$26.00 ©2011 IEEE

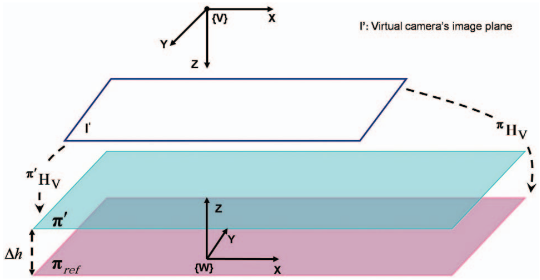

Figure 2. Extending homography for planes parallel to π_{ref}.

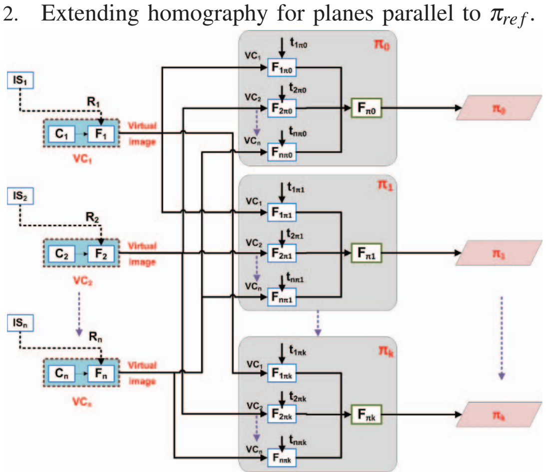

Figure 3. Parallel processing architecture.

in which $\mathbf{r}1$, $\mathbf{r}2$ and $\mathbf{r}3$ are the columns of the 3×3 rotation matrix and $\mathbf{t}$ is the translation vector between the π_{ref} and camera center [4]. We recall that all virtual cameras have the same rotation w.r.t world reference frame $\{W\}$. In other words one can think there is no rotation among the virtual cameras. Considering $\mathbf{t} = [\ t_1 \quad t_2 \quad t_3\ ]^T$ as the translation vector (among I' and π_{ref}) , the Eq. (3) can be rewritten as :

$$^{\pi}H_V^{-1} = \begin{bmatrix} f_x & 0 & f_x t_1 + u_0 t_3 \\ 0 & -f_y & f_y t_2 + v_0 t_3 \\ 0 & 0 & t_3 \end{bmatrix} \tag{4}$$

In order to estimate $\mathbf{t}$ an approach described in [3], [5] will be used. The homography matrix from virtual image plane to the world 3D plane π_{ref} has been already obtained as $^{\pi}H_V$ (Eq. (4)). For the sake of multi-layer reconstruction, it is desired to also obtain the homography matrix from a virtual image to another world 3D plane parallel to π_{ref} once we already have $^{\pi}H_V$ (see Fig. 2). Lets consider π' as a 3D plane which is parallel to π_{ref} and has a height Δh w.r.t it. Then by substituting t_3 in the equation (4) with $t_3 + \Delta h$ and expressing it via $^{\pi}H_V$ (the available homography between the virtual camera image plane and π_{ref}) we have:

$$^{\pi'}H_V^{-1}(\Delta h) = {}^{\pi}H_V^{-1} + \Delta h P \hat{\mathbf{k}}^T \tag{5}$$

where $P = [\ u_0 \quad v_0 \quad 1]^T$ is the principal point of the camera V and $\hat{\mathbf{k}}$ is the unit vector of the Z axis. Fig. 3 depicts the parallel processing architecture of the proposed approach.

III. Experiments

A real-time implementation of the 3D reconstruction algorithm has been developed as a prototype. The algorithm is developed using the C++ language, OpenCV library [2] and NVIDIA's CUDA software [1]. Fig. 4 shows the result of real-time 3D volumetric reconstruction using the proposed approach. The person is dressed in red and we used a mean-shift based segmentation algorithm. In this experiment 35 inertial-based virtual planes, with an interval distance of 5 *cm,* are used on order to register data for the sake of reconstruction. Then the 35 layers are stacked and visualized as can be seen in Fig. 4-right.

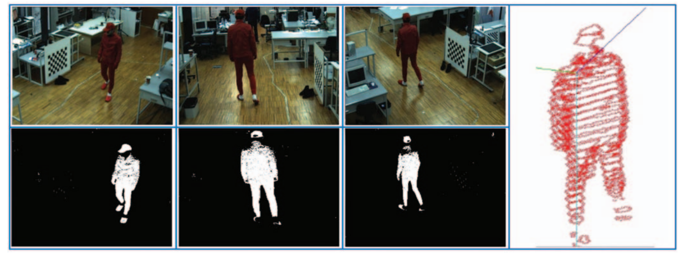

Figure 4. Results of the 3D volumetric reconstruction of the person in the scene. The space is observed by a IS-camera network. Left pictures show the images of the person before and after background subtraction. The right image depict the results of reconstruction. 35 inertial-based virtual planes, with an interval distance of 50 *mm,* are used for purpose of reconstruction.

IV. Conclusion and future work

3D data registration from 2D images have always been one of the very important computer vision topics. In this paper we presented a 3D volumetric reconstruction approach without using any planar ground assumption. The orientation data from IS is fused with visual information in order to make a virtual camera. IS is used also to define horizontal virtual planes in the scene. A real-time prototype is implemented based on the proposed algorithm. The implementation is carried out using a suitable parallel processing architecture adapted for GPU. As feature works: the idea is 1- To investigate the probabilistic fusion of range sensor in order to improve the registration, 2- To develop a multi-person tracking algorithm by taking advantage of the multi-layer registered data and 3- To conduct the work in the direction of cloud-robotics [6] where a robot is used as a service in a smart-room and has cooperation with the infrastructure sensors.

References

[1] Nvidia. http://www.nvidia.com/.
[2] Opencv. http://opencv.willowgarage.com/.
[3] Hadi Aliakbarpour and Jorge Dias. Human silhouette volume reconstruction using a gravity-based virtual camera network. In *Proceedings of the 13th International Conference on Information Fusion, 26-29 July 2010 EICC Edinburgh, UK*, 2010.
[4] Hadi Aliakbarpour and Jorge Dias. Imu-aided 3d reconstruction based on multiple virtual planes. In *DICTA'10 (the Australian Pattern Recognition and Computer Vision Society Conference), IEEE Pr., 1-3 December 2010, Sydney, Australia.*, 2010.
[5] Hadi Aliakbarpour and Jorge Dias. Inertial-visual fusion for camera network calibration. In *IEEE 9th International Conference on Industrial Informatics (INDIN 2011), July 2011.*, 2011.
[6] Yinong Chen, Zhihui Du, and M. Garci anda Acosta. Robot as a service in cloud computing. In *Service Oriented System Engineering (SOSE), 2010 Fifth IEEE International Symposium on*, pages 151–158, june 2010.
[7] M. Labrie and P. Hebert. Efficient camera motion and 3d recovery using an inertial sensor. In *Computer and Robot Vision, 2007. CRV '07. Fourth Canadian Conference on*, pages 55 –62, May 2007.