A Probabilistic Fusion Framework for 3D Reconstruction Using Heterogeneous Sensors

Hadi Aliakbarpour, João F. Ferreira, V. B. Surya Prasath, Kannappan Palaniappan, Guna Seetharaman, Jorge Dias

Abstract—**This letter proposes a framework to perform 3D reconstruction using a heterogeneous sensor network, with potential use in augmented reality (AR), human behavior understanding, smart-room implementations, robotics, and many other applications. We fuse orientation measurements from inertial sensors, images from cameras and depth data from Time of Flight (ToF) sensors within a probabilistic framework in a synergistic manner to obtain robust reconstructions. A fully probabilistic method is proposed to efficiently fuse the multi-modal data of the system.**

Index Terms—**Multi-modal fusion, heterogeneous sensor network, 3D reconstruction, probabilistic.**

I. Introduction

$\mathbf{I}$N the context of 3D reconstruction, camera networks are capable of providing multi-view images, with the further advantages of being passive sensors and of yielding additional information, such as surface color; however, 3D reconstruction using these sensors is sensitive to illumination, shadows, and homogeneous textures [1], [2]. On the other hand, Time-of-flight (ToF) depth sensors are able to provide depth information of a scene with much fewer degree of dependency on texture, however they do not provide color information. Therefore fusion of these disparate modalities in a synergic manner removes each of their individual shortcomings, while allowing for the overall harvesting of their advantages [3]. Moreover, using an appropriate probabilistic fusion approach would give a better result because of the inherent ability of probabilistic models in dealing with low-level heterogeneous information [4].

In this work, we propose a framework for volumetric 3D reconstruction using a network of heterogeneous sensors. A network of cameras, inertial sensors (known as IMU or IS), and ToF is considered to sense and gather information from the scene. Each camera is rigidly coupled to an IS. The 3D orientation provided by IS in each couple is used to define a virtual camera whose axis are aligned to the earth cardinal directions, and as a result, has a horizontal image plane. Using this 3D orientation, a set of virtual planes are defined in the scene (without any planar ground assumption) for the purpose of heterogeneous data registration. In order to fuse these heterogeneous data, a probabilistic fusion model is proposed. Experimental results showcase the practical advantages of the
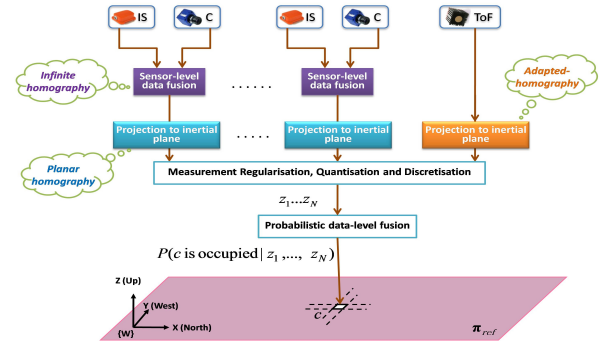
Fig. 1. Proposed approach uses sensor level fusion and probabilistic data level fusion.

geometrical solution of using inertial-based parallel planes to support independent 2D occupancy grids at sensor-level, and the robustness of the data-level probabilistic fusion model.

II. Proposed Approach

A. 3D Data Registration

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 to the earth fixed reference frame, see Fig. 1. In order to obtain image plane of virtual camera, a homography-based approach described in [5] has been used which fuses inertial data from IS and image plane of real camera to produce the corresponding virtual cameras image plane.

By taking the advantage of inertial data, a horizontal world plane π_{ref}, common between all virtual cameras, has been defined in the world reference frame $\{W\}$, see Fig. 2(a). 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, we do not use any real 3D plane inside the scene for estimating homography.

A 3D point $\mathbf{X} = [X\ Y\ Z\ 1]^T$ lying on π_{ref} is projected on virtual image plane as $\mathbf{x} = {}^{\pi_{ref}}H_v\,\mathbf{X}$, where ${}^{\pi_{ref}}H_v$ is a homography matrix that maps the π_{ref} to the virtual image plane,

$$ {}^{\pi_{ref}}H_v = \mathbf{K}\,[\ \mathbf{r}1\quad \mathbf{r}2\quad \mathbf{t}\ ] \tag{1} $$

in which $\mathbf{K}$ is the camera calibration matrix, $\mathbf{r}1$ and $\mathbf{r}2$ are the first and second columns of the 3×3 rotation matrix and $\mathbf{t}$ is the translation vector between π_{ref} and camera center [6].

B. Integrating Depth (ToF)

We utilize an occlusion criteria to register the depth data from ToF sensor onto the reference plane π_{ref}. Simple intersection of 3D points $^R\mathbf{X}$ from ToF and π_{ref} is not suitable since if we fuse this intersection ($^R\mathbf{X} \cap \pi_{ref}$) with the shadows created by cameras (using homography) then just the points on the boundary of the object (green arc in Fig. 2(b)) will remain as the result. We would like to keep the whole intersection between the object and the plane π_{ref}. Thus, we use the criterion that the point $\mathbf{x}$ will be mapped as "shadow" if and only if it lies on one of the range rays passing through the sensor's center and an object's surfacial point like $\mathbf{X}$. Otherwise it will be mapped as "not-shadow".

C. Probabilistic data fusion

Probabilistic data fusion from heterogeneous sensors [7] is utilized to compute the occupancy grid corresponding to reference plane π_{ref}, denoted as $\mathcal{Y}_{\pi_{ref}}$. If we let Z = [X, Y, S] denote complete sensor measurement vector of a sensor with $[S = 1]$ shadow, $[S = 0]$ not-shadow flag, and X, Y are the 2D coordinates from the sensor. The Bayesian sensor fusion modeling for N independent measurements by the sensor network at time t is given by,

$$P(Z_1^t \ldots Z_N^t, O_c^t) = P(O_c^t) \prod_{i=1}^{N} P(Z_i^t | O_c^t), \qquad (2)$$

where O_c (i.e. the occupancy O_c of each independent cell $c \in \mathcal{Y}_{\pi_{ref}}$) is a binary variable signaling the occupancy of cell c, with $[O_c = 1]$ - occupied, and $[O_c = 0]$ - empty. The probability of cell c being occupied, $P([O_c^t = 1]|z_1^t \ldots z_N^t)$, can then be inferred, assuming that the prior knowledge on occupancy is given by the cell's state in the previous time instant, $P(O_c^t) \equiv P(O_c^{t-1}|z_1^{t-1} \ldots z_N^{t-1})$ (i.e. a simple temporal Bayesian filter update). Therefore, applying Bayes' rule and marginalization to Eqn. (2) , and denoting $o_c^t \equiv [O_c^t = 1]$ and $\bar{o}_c^t \equiv [O_c^t = 0]$, we obtain

$$
\begin{aligned}
P(o_c^t|z_1^t \ldots z_N^t) &= \frac{P(o_c^t) \prod_{i=1}^{N} P(z_i^t|o_c^t)}{P(z_1^t \ldots z_N^t)} \\
&= \frac{P(o_c^t) \prod_{i=1}^{N} P(z_i^t|o_c^t)}{P(o_c^t) \prod_{i=1}^{N} P(z_i^t|o_c^t) + P(\bar{o}_c^t) \prod_{i=1}^{N} P(z_i^t|\bar{o}_c^t)},
\end{aligned}
\qquad (3)
$$

which can be computed analytically, using an efficient, closed-form expression. Repeating this operation for all cells constituting $\mathcal{Y}_{\pi_{ref}}$, the full state O^t of the occupancy grid can be estimated for a particular time instant t.

III. Experimental Results

We utilized synchronized AVT Prosilica GC650C GigE Color cameras with rigidly coupled Xsens MTx IS. The IS provided the 3D orientation to obtain the virtual camera, and horizontal planes. Microsoft Kinect sensor was used as a ToF sensor. An efficient GPU-CUDA implementation provides real-time 3D reconstructions. Figure 3 shows an example output obtained by our system of a dynamically moving person. Totally, 57 inertial-based virtual planes were used for registering 3D data on the scene using the homography

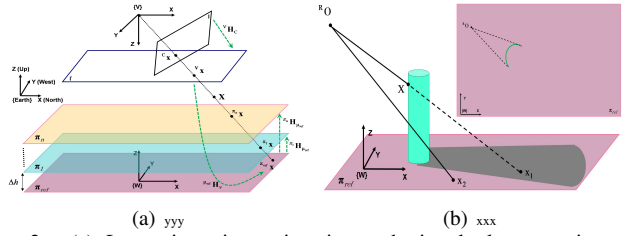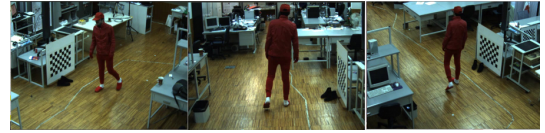(a) yyy   (b) xxx

Fig. 2.   (a) Image intensity registration and virtual plane creation with homographies. (b) A ToF observing a cylinder. The intention is to register the intersection of the object and the reference plane π_{ref} using homography.

(a) Intensity images

(b) Before depth fusion   (c) ToF Depth   (d) After depth fusion

Fig. 3.   The raw data recorded by sensors network (a and c) and the 3D reconstruction output (b and d).

method, and the geometrical solution obtained provides 2D occupancy grids. The methodology is also a natural and efficient way to integrate data yielded by heterogeneous sensors, by intrinsically and explicitly taking into account the uncertainty of measurements provided by each sensor. Further, it offers elegant solution via probabilistic fusion for the fusing data from a network of heterogeneous sensors.

IV. Conclusions

We considered a geometric framework for volumetric 3D reconstruction using a network of heterogeneous sensors. A network of cameras, inertial and ToF sensors were used to derive local depth maps. A data-level probabilistic fusion was performed to efficiently fuse the heterogeneous data with the ability to take into account the uncertainty of each measurement. The framework proposed here has applications in human behavior understanding, tracking, human-robot-interaction etc.

References

[1] T. C. S. Azevedo, J. M. R. S. Tavares, M. A. P. Vaz, "3D object reconstruction from uncalibrated images using an off-the- shelf camera," Adv. Comput. Vis. Med. Imag. Proc., vol. 13, pp. 117–136, 2009.
[2] M. A. Brodie, A. Walmsley, W. Page, "The static accuracy and calibration of inertial measurement units for 3D orientation," Comp. Meth. Biomech. Biomed. Eng., vol. 11, pp. 641–648, 2008.
[3] R. C. Luo, C.-C. Yih, K. L. Su, "Multisensor fusion and integration: approaches, applications, and future research directions," IEEE Sensors, vol. 2, pp. 107–119, 2002.
[4] A. Elfes, "Multi-source spatial data fusion using Bayesian reasoning," Data Fusion in Robotics and Machine Intelligence, Academic Press, 1992.
[5] H. Aliakbarpour, V. B. S. Prasath, K. Palaniappan, G. Seetharaman, J. Dias, "Heterogeneous multi-view information fusion: Review of 3-D reconstruction methods and a new registration with uncertainty modeling," IEEE Access, vol. 4, no. 1, pp. 8264–8285, 2016.
[6] R. Hartley, A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge Univ. Press, 2004.
[7] J.F. Ferreira, J. Dias. "Probabilistic Approaches for Robotic Perception," Springer, 2014.