Human Silhouette Volume Reconstruction Using a Gravity-based Virtual Camera Network

Hadi Aliakbarpour
Institute of Systems and Robotics
Elect. and Comp. Engineering Department
University of Coimbra.
hadi@isr.uc.pt

Jorge Dias
Institute of Systems and Robotics
Elect. and Comp. Engineering Department
University of Coimbra.
jorge@isr.uc.pt

Abstract – *The article represents a method to perform the* Shape From Silhouette *(SFS) of human, based on gravity sensing. A network of cameras is used to observe the scene. The extrinsic parameters among the cameras are initially unknown. An IMU is rigidly coupled to each camera in order to provide gravity and magnetic data. By applying a data fusion between each camera and its coupled IMU, it becomes possible to consider a downward-looking virtual camera for each camera within the network. Then extrinsic parameters among virtual cameras are estimated using the heights of two 3D points with respect to one camera within the network. Registered 2D points on the image plane of each camera is reprojected to its virtual camera image plane, using the concept of* infinite homography. *Such a virtual image plane is horizontal with a normal parallel to the gravity. The 2D points from the virtual image planes are back-projected onto the 3D space in order to make conic volumes of the observed object. From intersection of the created conic volumes from all cameras, the silhouette volume of the object is obtained. The experimental results validate both feasibility and effectiveness of the proposed method.*

Keywords: Shape From Silhouette (SFS), IMU (Inertial Measurement Unit), 3D reconstruction, virtual camera, virtual image plane, infinite homography.

1 introduction

3D silhouette reconstruction of human is highly useful for many applications including human behaviour understanding. SFS is a known method to automatically reconstruct 3D shape of an object by using a set of images which are taken from multiple views. In this method the position and orientation of the cameras need to be known. It means that the camera network has to be initially calibrated. There are several methods to calibrate camera network[25, 3, 4, 12, 30]. However they work on either by moving a bright spot object (such as LED) through the darkened scene, placing some calibration pattern in different orientation, or using vanishing points of some structure in the scene. In some cases, it is not possible to make the scene dark enough or either difficult to move an object inside the scene. Therefore most of the extrinsic calibration approaches can not easily be applied for some cases. In this paper a method to perform the SFS, based on making a virtual camera network, is proposed. A network of cameras is used to observe the object within the scene. An IMU is rigidly coupled to each camera in order to provide gravity and magnetic data. By applying a data fusion between each camera and its coupled IMU, it becomes possible to consider a downward-looking virtual camera for each camera within the network. Then extrinsic parameters among virtual cameras are estimated using the heights of two 3D points with respect to one camera within the network. Registered 2D points on the image plane of each camera is reprojected to its virtual camera image plane, using the concept of *infinite homography*. Such a virtual image plane is horizontal with a normal parallel to the gravity. The 2D points from the virtual image planes are back-projected onto the 3D space in order to make conic volumes of the observed object. From intersection of the created conic volumes from all cameras, the silhouette volume of the object is obtained. Based on the proposed method even just one IMU-camera couple can be used to make a camera network by moving it inside the scene. The only prerequisite of the scene structure is to know the heights of two arbitrary 3D points with respect to one camera within the network and there is no more need of doing any in-scene calibration operation.

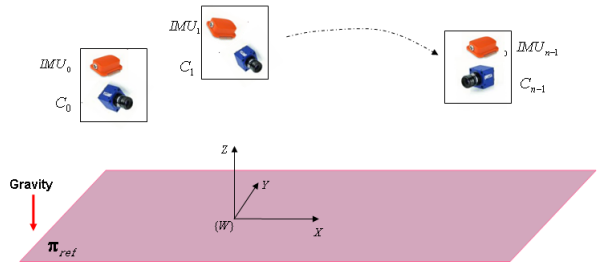

Figure 1: A network of IMU-camera couples and a horizontal plane

1.1 Previous work

In [26], Wada presented a homography-based parallel intersection method to reconstruct object's volume. Khan in [17] proposed a homographic framework for the fusion of multi-view silhouettes. A marker-less 3D human motion capturing approach is introduced in [20] using multiple views. Michoud in [21] investigated a multi-view framework to compute a 3D shape estimation of multiple objects from silhouette and without ghost object. Takahashi in [16, 14, 15] proposed some remarks on 3D human body posture estimation extracted from multi cameras using silhouette volume intersection technique. Zhang in [29] introduced an algorithm for 3D projective reconstruction based on infinite homography. In his approach he improved the estimation of homography matrix in such a way that instead of having 4 points on a reference plane it needs 3 points. An octree-based fusion of shape from silhouette and shape from structured light is proposed by Kampel in [13]. Lai and Yilmaz in [18] used images from uncalibrated cameras for performing projective reconstruction of buildings based on SFS approach where buildings structure is used to compute vanishing points. Chen and Chai in [7] proposed a method to perform 3D reconstruction of human motion and skeleton from uncalibrated monocular video. In [27] a multi-camera network system is applied for markerless 3D human body reconstruction. Zhang and Hanson in [31] implemented a 3D Reconstruction based on homography mapping. Franco in [10] used a Bayesian occupancy grid to represent the silhouette cues of objects. The use of IMU sensor to accompany compute vision applications is recently attracting attentions of the researchers. Dias in [8] investigated the cooperation between visual and inertial information. Lobo and Dias[19] proposed an efficient method to estimate the relative pose of a camera and an IMU. In our previouse works [2, 1], IMU was used in order to perform calibration between a 3D laser range finder and a stereo camera as well as data registration. Moreover the method presented in [2] was extented in order to calibrate a stereo camera and a 3D tracker[9]. Mirisola in [23] used a rotation-compensated imagery for the aim of trajectory by aiding inertial data. Fusion of image and inertial data is also investigated by Bleser [5] for the sake of tracking in the mobile augmented reality.

1.2 Outline

As mentioned, in this paper a method to perform the SFS, based on making a virtual camera network, is proposed. The data from each coupled camera and IMU are fused in order to actualize a virtual downward looking camera. After that a novel 2-known-heights based method is proposed to estimate the extrinsic parameters between cameras within the virtual camera network. Then the registered images by the virtual camera network are used to perform the SFS method for the sake of human silhouette volume reconstruction. The use of gravity and magnetic data to perform SFS-based 3D reconstruction is the main contribution of this paper. The article is structured as follows: The models of camera and IMU are explained in Sec. 2. The method to actualize a network of virtual cameras is proposed in Sec. 3. In Sec. 4, an algorithm to carry out the SFS-based reconstruction is suggested. Sec. 5 is dedicated to the experiments and eventually the conclusion and future work is presented in Sec. 6.

2 Sensor model

In a pinhole camera model, a 3D point $\mathbf{X} = [\ X \quad Y \quad Z \quad 1\ ]^T$ in the scene and its corresponding projection $\mathbf{x} = [\ x \quad y \quad 1\ ]^T$ (both $\mathbf{X}$ and $\mathbf{x}$ are expressed in normalized homogeneous form) are related via a 3×4 matrix P (called *Projection matrix*) through the following equation [11]:

$$x = PX \tag{1}$$

$$P = K\,[R|t] \tag{2}$$

where K is the *camera calibration matrix, R* is rotation matrix between world and camera coordinate systems and t is translation vector between world and camera coordinate systems which is equal to $t = -RC$, C being is the center of the camera expressed in the world coordinate system. The camera matrix K, which is also called *intrinsic parameter matrix,* is defined by [11]:

$$K = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3}$$

in which f_x and f_y represent the focal length of the camera (in terms of pixel scale) in the directions of x and y. The u_0 and v_0 are the elements of the principal point vector p [11]. In order to map points from one plane to another plane the concept of Homography [11] is used. Lets consider a 3D plane is observed by two cameras with $P = [I|0]$ and $P^{'} = [R|t]$ (concerning first camera center as world reference frame). Also assume that x_1 and x_2 are the image points of a 3D point X lying on the 3D plane. Then x_1 and x_2 are called a pair of correspondence points and the relation between them can be expressed as $x_2 = H\,x_1$ in which H is a 3×3 matrix called *planar homography* induced by the 3D plane [28] and is equal to (up to scale)

$$H = R + \frac{1}{d}\,t\,n^T \tag{4}$$

in which R and t are rotation matrix and translation vector between the two cameras centers, n is Normal of the 3D plane and d is the orthogonal distance between the 3D plane and camera center. Applying the camera calibration matrices K and $K^{'}$ and consequently having $P = K\,[I|0]$ and $P^{'} = K^{'}\,[R|t]$ as camera projection matrices then corresponding equation will become [11]:

$$H = K^{'}\left(R + \frac{1}{d} t\, n^{T}\right) K^{-1} \tag{5}$$

Regarding the IMU, just its the gravity and magnetic data are used in this work. The complete model of this sensor is described in [19].

3 Virtual camera network construction

3.1 Reference Frames

Fig. 1 shows a setup with n couples of IMU-camera and a horizontal plane. In this setup each camera is rigidly coupled to an IMU. If the scene is static, then even a single IMU-camera is sufficient, since it can be put in different places of the scene (in this case, in any position the data from IMU-camera couple should be stored and after that it can be replaced to a new position). Using the orientation given by IMU it becomes possible to assume a virtual camera beside of each real camera within the network. In order to calibrate the network, four different reference frames are involved (see Fig. 2):

- *Real camera reference frame* $\{C\}$: The local coordinate system of a camera C is expressed as $\{C\}$.
- *Earth reference frame* $\{E\}$: Which is an earth fixed reference frame having its X axis in the direction of *North*, Y in the direction of *West* and Z upward.
- *Inertial Measuring Unit local reference frame* $\{IMU\}$: This is the local reference frame of an IMU sensor which is defined w.r.t. to the earth reference frame $\{E\}$.
- *Virtual camera reference frame* $\{V\}$: As explained, for each real camera C, a virtual camera V, is considered by the aid of a rigidly coupled IMU to that. $\{V\}$ indicates the reference frame of such a virtual camera. The centers of $\{C\}$ and $\{V\}$ coincide and therefore there is just a rotation among these two reference frames.

3.2 Rotation compensation

The idea here is to introduce a method to recover the rotation among cameras within the network. Then it will be possible to compensate rotation for each camera or in the other words having a camera network with no rotation among them. A camera network setup is shown in Fig. 1. We start to explain the method for one camera and then it can be extended for all cameras. Fig. 3 shows the center, optical axis, image plane and principal points of a real camera C and its corresponding virtual camera V. Here the image plane of real camera and virtual camera are named as I and $I^{'}$, respectively. Based on our definition V is downward looking camera and has optical axis parallel to the gravity vector. Thus $I^{'}$ becomes

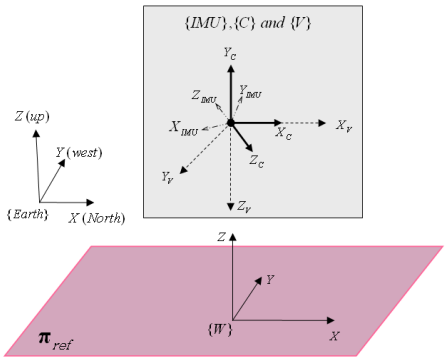

Figure 2: Involved reference frames in the proposed approach and a horizontal world plane. The plane can be any world plane which has a normal parallel to the gravity

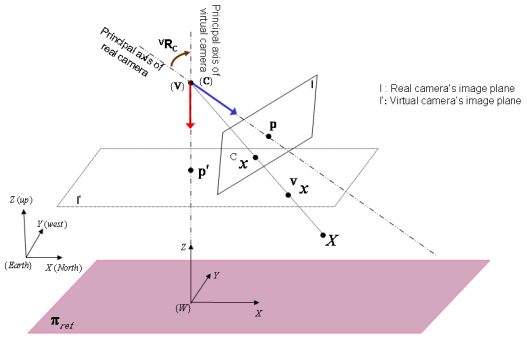

Figure 3: Virtual Camera Using IMU-aided Homography

a horizontal image plane at a distance f below the camera sensor, f being the focal length [22].

In this fashion the intention is to register a 3D point X onto $I^{'}$. This can be done in two steps. Firstly X is registered onto I as ${}^{c}x = K\,[I|0]\,X$. Then the 2D point ${}^{c}x$ can be reprojected onto $I^{'}$ as follows:

$$^{v}x = {}^{v}H_{c}\ {}^{c}x \tag{6}$$

in which ${}^{v}H_{c}$ is a 3×3 homography matrix between I and $I^{'}$. As described before, the real camera C and virtual camera V have their centers coincided to each other, so the transformation between these two cameras can be expressed just by a rotation matrix (see Fig. 3). In this case ${}^{v}H_{c}$ is called *infinite homography* since there is just a pure rotation between real camera and virtual camera centers [11]. Such an infinite homography can be computed using a limiting process on Eq. (5) by considering either $d \to \infty$ or $t \to 0$:

$$^{V}H_{C} = \lim_{d\to\infty} K\left({}^{V}R_{C} + \frac{1}{d} t\, n^{T}\right) K^{-1} = K\, {}^{V}R_{C}\, K^{-1} \tag{7}$$

where ${}^{V}R_{C}$ is the rotation matrix between $\{C\}$ and $\{V\}$ [24]. The rotation matrix ${}^{V}R_{C}$ can be computed through three consecutive rotations (see the reference frames in Fig. 2) as follows:

$$^{V}R_C = {}^{V}R_E \; {}^{E}R_{IMU} \; {}^{IMU}R_C \tag{8}$$

First one ${}^{IMU}R_C$ is to transform from real camera reference $\{C\}$ to the IMU local coordinate $\{IMU\}$. The second one ${}^{E}R_{IMU}$ transforms from the $\{IMU\}$ to the earth fixed reference $\{E\}$ and the last one ${}^{V}R_E$ is to transform from $\{E\}$ to virtual camera reference frame $\{V\}$. Here we continue to explain how to compute these three consecutive rotation matrices. In order to estimate the rotation between camera and IMU (${}^{IMU}R_C$) , *Camera Inertial Calibration Toolbox* [19] is used which is a toolbox to calibrate a rigid couple of a IMU and camera. Rotation from IMU to earth ,${}^{E}R_{IMU}$, is given by the IMU sensor w.r.t $\{E\}$. Since the $\{E\}$ has the Z upward but the virtual camera is supposed to be downward-looking (with a downward Z) then the following rotation is applied to reach to the virtual camera reference frame:

$$^{V}R_E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{9}$$

Having ${}^{V}R_C$ for each cameras and applying Eq. 7 for any camera within the network, we will have a set of parallel (and horizontal) virtual image planes in such a way that there is no rotation among them. This means that by now we have solved the problem of rotation recovery for cameras. In the next section we proceed to explain a method to recover the translations among the cameras in the network.

3.3 Translation among virtual cameras

In previous sub-section, it was explained how to reach to a network of virtual cameras such a way that there is no rotation among them. It means that we have reduced the problem of calibration to just a *"pure translation*" case. This section is dedicated to propose a method to estimate the translation between the mentioned virtual cameras. Obviously since the center of each virtual camera is coincided to its corresponding real camera then the translations between virtual cameras set are equal to the real ones. As described before, the only requirement from the scene is to have the heights of two arbitrary 3D points such $\mathbf{X}_1 = [\ X_1 \quad Y_1 \quad Z_1\ ]^T$ and $\mathbf{X}_2 = [\ X_2 \quad Y_2 \quad Z_2\ ]^T$ (see Fig. 4) w.r.t one camera (namely V_0) within the network.

Suppose ${}^{0}\mathbf{X}_1 = [\ {}^{0}X_1 \quad {}^{0}Y_1 \quad {}^{0}Z_1\ ]^T$ and ${}^{0}\mathbf{X}_2 = [\ {}^{0}X_2 \quad {}^{0}Y_2 \quad {}^{0}Z_2\ ]^T$ are coordinations of the two 3D points $\mathbf{X}_1$ and $\mathbf{X}_2$ expressed in the first virtual camera center, respectively. Based on the assumption, the parameters ${}^{0}Z_1$ and ${}^{0}Z_2$ which indicate the heights of $\mathbf{X}_1$ and $\mathbf{X}_2$ in $\{V_0\}$ are known. Therefore the term "*height*" here is equal to the Z component of the 3D point. Then using projective property of a camera we can have all three components of ${}^{0}\mathbf{X}_1$ and ${}^{0}\mathbf{X}_2$ numerically computed in a metric scale using the Eq. (10):

$$\begin{cases} {}^{0}\mathbf{X}_1 = {}^{0}Z_1 \, (\mathbf{K}_1^{-1} \, {}^{0}\mathbf{x}_1) \\ {}^{0}\mathbf{X}_2 = {}^{0}Z_2 \, (\mathbf{K}_1^{-1} \, {}^{0}\mathbf{x}_2) \end{cases} \tag{10}$$

where ${}^{0}\mathbf{x}_1$ and ${}^{0}\mathbf{x}_2$ are respectively the imaged points of $\mathbf{X}_1$ and $\mathbf{X}_2$ in the first virtual camera image plane. The same can be considered for the second virtual camera. Suppose ${}^{1}\mathbf{X}_1 = [\ {}^{1}X_1 \quad {}^{1}Y_1 \quad {}^{1}Z_1\ ]^T$ and ${}^{1}\mathbf{X}_2 = [\ {}^{1}X_2 \quad {}^{1}Y_2 \quad {}^{1}Z_2\ ]^T$ are respectively coordinations of the 3D points $\mathbf{X}_1$ and $\mathbf{X}_2$ expressed in the second virtual camera center ($\{V_1\}$). Then likewise using projective property of a camera we can have the following equation:

$$\begin{cases} {}^{1}\mathbf{X}_1 = {}^{1}Z_1 \, (\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_1) \\ {}^{1}\mathbf{X}_2 = {}^{1}Z_2 \, (\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_2) \end{cases} \tag{11}$$

In contrary to the Eq. (10), Eq. (11) can not be numerically computed yet, since it has two unknown values for ${}^{1}Z_1$ and ${}^{1}Z_2$ (the heights of the 3D points w.r.t $\{V_1\}$). The terms $(\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_1)$ and $(\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_2)$ in Eq. (11) as well express the 3D position of the points ${}^{1}\mathbf{X}_1$ and ${}^{1}\mathbf{X}_2$ however up to scale factors ${}^{1}Z_1$ and ${}^{1}Z_2$. Here it is desirable to rewrite the Eq. (11) as the following:

$$\begin{cases} {}^{1}\mathbf{X}_1 = {}^{1}Z_1 \, {}^{1}\hat{\mathbf{X}}_1 \\ {}^{1}\mathbf{X}_2 = {}^{1}Z_2 \, {}^{1}\hat{\mathbf{X}}_2 \end{cases} \tag{12}$$

where ${}^{1}\hat{\mathbf{X}}_1 = (\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_1)$ and ${}^{1}\hat{\mathbf{X}}_2 = (\mathbf{K}_2^{-1} \, {}^{1}\mathbf{x}_2)$. Then the Eq. (10) and Eq. (12) can be related through the translation vector between $\{V_0\}$ and $\{V_1\}$ as:

$$\begin{cases} {}^{0}\mathbf{X}_1 = {}^{1}\mathbf{X}_1 + \mathbf{t} & = {}^{1}Z_1 \, {}^{1}\hat{\mathbf{X}}_1 + \mathbf{t} \\ {}^{0}\mathbf{X}_2 = {}^{1}\mathbf{X}_2 + \mathbf{t} & = {}^{1}Z_2 \, {}^{1}\hat{\mathbf{X}}_2 + \mathbf{t} \end{cases} \tag{13}$$

where $\mathbf{t} = (\ t_1 \quad t_2 \quad t_3\ )^T$. In Eq. (13) there are five unknown parameters, ${}^{1}Z_1$, ${}^{1}Z_2$, t_1, t_2, t_3, and six linear equations. In order to estimate the five unknowns Eq. (13) can be arranged in the form of

$$\mathbf{A}\,\mathbf{x} = \mathbf{B} \tag{14}$$

where

$$\mathbf{A} = \begin{bmatrix} {}^{1}\hat{\mathbf{X}}_1 & \mathbf{0}_{3\times1} & \mathbf{I}_{3\times3} \\ \mathbf{0}_{3\times1} & {}^{1}\hat{\mathbf{X}}_2 & \mathbf{I}_{3\times3} \end{bmatrix},$$

$$\mathbf{x} = [\ {}^{1}Z_1 \quad {}^{1}Z_2 \quad t_1 \quad t_2 \quad t_3\ ]^T, \mathbf{B} = \begin{bmatrix} {}^{0}\mathbf{X}_1 \\ {}^{0}\mathbf{X}_2 \end{bmatrix}$$

$\mathbf{x}$ in Eq. (14) can thus be estimated using the least square approach and consequently the translation vector between the two virtual cameras' frames, $\{V_0\}$ and $\{V_1\}$, is estimated. In the same way, the translations between all other virtual cameras and $\{V_0\}$ can be estimated.

4 Silhouette volume reconstruction

In this section, the applied SFS method for the aim of 3D reconstruction is explained. As introduced before, a network of virtual cameras is used to observe the object. Here having

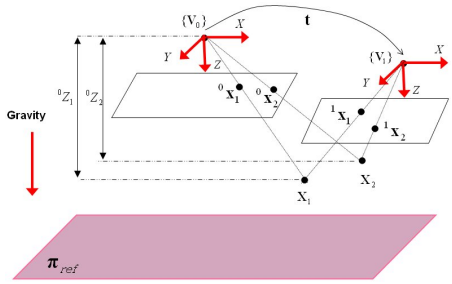

Figure 4: Translation between two virtual cameras (corresponding to two real cameras: first one from the structure camera and the other one here is on a mobile robot).

a sufficient background subtraction algorithm is an assumption. Moreover, the only interesting feature from the object in this approach is its silhouette. The silhouette of the object is registered on the image planes of the cameras in the network. Then 2D points of the image plane of each camera are re-projected onto its corresponding virtual camera's image plane ($I^{'}$), based on the proposed method. Then the 2D points from the virtual image planes are back-projected onto the 3D space in order to make conic volumes of the observed object. Using intersection of the created conic volumes from all cameras, the silhouette volume of the object is obtained. In order to do these operations the following algorithm is proposed (Alg. 1). Here, $\{voxel\}$ is the set of voxel involved in the 3D space, $\{camera\}$ is the set of cameras. K_c is the camera matrix corresponding to c. In the case of using just one camera (by moving it through the scene) then there will be just one K for all cameras. t_v is the translation vector for the virtual camera v which can be calculated by the proposed method in 3.3. $I^{'}(\mathbf{x})$ indicate the intensity of the virtual image $I^{'}$ in the cell $\mathbf{x}$.

Algorithm 1 3D Reconstruction

for each *vox* in $\{voxel\}$ **begin**
  vox:= '*occupied*'
  consider X as 3D position of v in the space.
  for each c in $\{camera\}$ **begin**
    consider v as the virtual camera of c
    consider $I^{'}$ as the image plane of v
    compute $\mathbf{x} = K_c(X + t_v)$
    if $I^{'}(\mathbf{x}) < thresh$ then
      vox:='*empty*'
    endif
  end
end

5 Experiments

An experiment is performed based on the proposed approach. Fig. 5 show a rigid Camera-IMU couple which is used in our experiment. In this experiment just one Camera-IMU couple is used and the camera network is made by

Figure 5: Camera-IMU setup

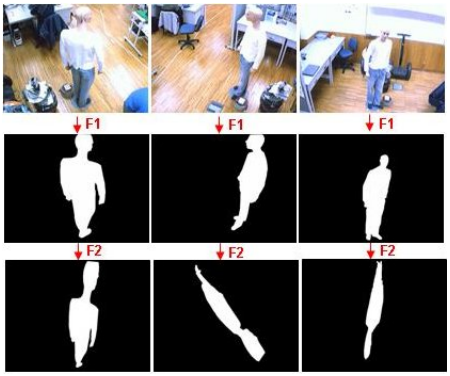

Figure 6: From real image planes to virtual image planes: F1: Background subtraction process. F2: Image plane of virtual camera.

manually placing it in different position. The camera is a simple 640×480 *FireWire Unibrain* camera. A *MTi-Xsens* is used as the IMU sensor. Firstly the intrinsic parameters of the camera camera is estimated using Bouguet Camera Calibration Toolbox[6] :

$$K = \begin{bmatrix} 750.9819 & 0 & 367.5754 \\ 0 & 751.8286 & 292.6940 \\ 0 & 0 & 1 \end{bmatrix} \quad (15)$$

Then *Camera Inertial Calibration Toolbox* [19] is used for the sake of extrinsic calibration between the camera and IMU:

$${}^{C}R_{IMU} = \begin{bmatrix} 0.0032 & -0.9996 & -0.0286 \\ 0.0179 & 0.0286 & -0.9994 \\ 0.9998 & 0.0027 & 0.0179 \end{bmatrix} \quad (16)$$

The Camera-IMU couple was placed in three different positions. In each position an image and also the gravity-magnetic data of the IMU is recorded. Fig. 6, the first row, shows the image planes of the real cameras. Then a virtual camera for each real camera is constructed. Further, the registered 2D-points on the image plane are reprojected onto the virtual image planes (see Fig. 6-third row). Translations among cameras are calculated by using the heights of two points in the scene. As the next step, the 3D reconstruction of the maniken is obtained usign the proposed algorithm (Alg. 1). Fig. 7 represents the result (from three different views) after aplying the algorithm on the data-set shown in Fig. 6. The used resolution for the volume is as $20mm$ in each axis.

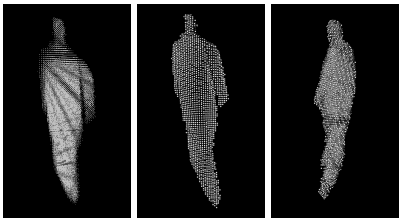

Figure 7: The result of the proposed method for the SFS-based 3D reconstruction: These three figures are taken from three different views. It is the result of Alg. 1, applied to the data-set shown in Fig. 6.

6 Conclusion and future work

In this paper the Shape From Silhouette (SFS) method is applied for 3D reconstruction of human using a network of sensors. The network comprises of IMUs and cameras. The camera network is calibrated just based on gravity and magnetometer data and measuring the height of two 3D points in the scene w.r.t one camera within the network. The need for human operation for camera network calibration is minimum. Using the proposed method it is possible to rapidly mount the IMU-Camera couples in the scene, measuring the heights of two 3D point and then starting to monitor the scene. As future work the intention is to continue this work for surveillance applications where there are more modalities such as range and vocal data.

7 Acknowledgment

Hadi Ali Akbarpour is supported by the FCT (Portuguese Fundation for Science and Technology). This work is partially supported by the European Union within the FP7 Project PROMETHEUS, www.prometheus-FP7.eu. The authors would like to thank *Amilcar Ferreira* and *Hugo Faria* for their helps in the data collection phase.

References

[1] Hadi Aliakbarpour, J. F. Ferreira, K. Khoshhal, and Jorge Dias. A novel framework for data registration and data fusion in presence of multi-modal sensors. In *in Proceedings of the DoCEIS 2010- Doctoral Conference on Computing, Electrical and Industrial Systems - Lisbon, Portugal, Feb. 22-24, 2010, Published by Springer Boston, IFIP Advances in Information and Communication Technology (Emerging Trends in Technological Innovation).*, volume 314/2010, pages 308–315, 2010.

[2] Hadi Aliakbarpour, Pedro Nunez, Jose Prado, Kamrad Khoshhal, and Jorge Dias. An efficient algorithm for extrinsic calibration between a 3d laser range finder and a stereo camera for surveillance. In *14th International Conference on Advanced Robotics (ICAR 2009)*, 2009.

[3] Joao P. Barreto and Kostas Daniilidis. Wide area multiple camera calibration and estimation of radial distortion. In *Int. Work. on Omnidirectional Vision, Camera Networks and Non-Classical Cameras. Prague, May 2004.*, 2004.

[4] Silvain Beriault, Pierre Payeur, and Gilles Comeau. Flexible multi-camera network calibration for human gesture monitoring. In *ROSE 2007 - IEEE International Workshop on Robotic and Sensors Environments, Ottawa - Canada, 12-13 October 2007.*, 2007.

[5] Bleser, Wohlleber, Becker, and Stricker. Fast and stable tracking for ar fusing video and inertial sensor data. pages 109–115. Short Papers Proceedings. Plzen: University of West Bohemia, 2006.

[6] Jean-Yves Bouguet. Camera calibration toolbox for matlab. In *www.vision.caltech.edu/bouguetj*.

[7] Yen-Lin Chen and Jinxiang Chai. 3d reconstruction of human motion and skeleton from uncalibrated monocular video. In *The Ninth Asian Conference on Computer Vision (ACCV 2009)*, 2009.

[8] Jorge Dias, Jorge Lobo, and Luis A. Almeida. Cooperation between visual and inertial information for 3d vision. In *Proceedings of the 10th Mediterranean Conference on Control and Automation - MED2002 Lisbon, Portugal, July 9-12, 2002.*, 2002.

[9] Diego R. Faria, Hadi Aliakbarpour, and Jorge Dias. Grasping movements recognition in 3d space using a bayesian approach. In *14th International Conference on Advanced Robotics (ICAR 2009)*, 2009.

[10] Jean-Sebastien Franco and Edmond Boyer. Fusion of multi-view silhouette cues using a space occupancy grid. In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV05)*, 2005.

[11] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. CAMBRIDGE UNIVERSITY PRESS, 2003.

[12] B.W. He and Y.F. Li. Camera calibration from vanishing points in a vision system. *Optics & Laser Technology, Elsevier.*, 40:555–561, 2007.

[13] M. Kampel, S. Tosovic, and R. Sablatnig. Octree-based fusion of shape from silhouette and shape from structured light. In *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on, IEEE*, 2002.

[14] Masafumi Hashimoto Kazuhiko Takahashi, Yusuke Nagasawa. Remarks on 3d human body's feature extraction from voxel reconstruction of human body posture. pages 121–126. Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics, December 15 -18, 2007, Sanya, China., 2007.

[15] Masafumi Hashimoto Kazuhiko Takahashi, Yusuke Nagasawa. Remarks on real-time 3d human body posture estimation using multi-camera system. pages 2360–2365. The 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON) Nov. 5-8, 2007, Taipei, Taiwan, 2007.

[16] Yusuke Nagasawa Kazuhiko Takahashi and Masafumi Hashimoto. Remarks on 3d human body posture estimation system using sinple multi-camera system. IEEE, 2006.

[17] Saad M. Khan, Pingkun Yan, and Mubarak Shah. A homographic framework for the fusion of multi-view silhouettes. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007.

[18] Po-Lun Lai and Alper Yilmaz. Projective reconstruction of building shape from silhouette images acquired from uncalibrated cameras. In *ISPRS Congress Beijing 2008, Proceedings of Commission III*, 2008.

[19] Jorge Lobo and Jorge Dias. Relative pose calibration between visual and inertial sensors. *International Journal of Robotics Research, Special Issue 2nd Workshop on Integration of Vision and Inertial Sensors*, 26:561–575, 2007.

[20] Brice Michoud, Erwan Guillou, and Sada Bouakaz. Real-time and markerless 3d human motion capture using multiple views. *Human Motion-Understanding, Modeling, Capture and Animation, Springer Berlin/Heidelberg.*, 4814/2007:88–103, 2007.

[21] Brice Michoud, Bouakaz Saida, Guillou Erwan, and Briceno Hector. Largest silhouette-equivalent volume for 3d shapes modeling without ghost object. In *M2SFA2 2008: Workshop on Multi-camera and Multimodal Sensor Fusion, Marseille, France.*, 2008.

[22] Luiz G. B. Mirisola and Jorge Dias. Tracking from a moving camera with attitude estimates. In *ICR08*, 2008.

[23] Luiz G. B. Mirisola, Jorge Dias, and A. Traca de Almeida. Trajectory recovery and 3d mapping from rotation-compensated imagery for an airship. In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems San Diego, CA, USA, Oct 29 - Nov 2, 2007*, 2007.

[24] Luiz Gustavo Bizarro Mirisola. *Exploiting attitude sensing in vision-based navigation, mapping and tracking including results from an airship*. PhD thesis, 2009.

[25] Tomas Svoboda, Daniel Martinec, and Tomas Pajdla. A convenient multi-camera self-calibration for virtual environments. *PRESENCE: Teleoperators and Virtual Environments, MIT Press.*, 14:407–422, 2005.

[26] Toshikazu Wada, Xiaojun Wu, Shogo Tokai, and Takashi Matsuyama. Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras. In *Computer Architectures for Machine Perception, 2000. Proceedings. Fifth IEEE International Workshop on 11-13 Sept. 2000*, pages 331–339, 2000.

[27] Tao Yang, Yanning Zhang, Meng Li, Dapei Shao, and Xingong Zhang. A multi-camera network system for markerless 3d human body voxel reconstruction. In *Fifth International Conference on Image and Graphics, 2009. ICIG '09.*, 2009.

[28] Jana Kosecka Yi Ma, Stefano Soatta and S. Shankar Sastry. *An invitation to 3D vision*. Springer, 2004.

[29] Quan-Bing Zhang, Hai-Xian Wang, and Sui Wei. A new algorithm for 3d projective reconstruction based on infinite homography. In *Machine Learning and Cybernetics, 2003 International Conference on, IEEE*, 2003.

[30] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999.

[31] Zhongfei Zhang and Allen R. Hanson. 3d reconstruction based on homography mapping. In *In ARPA Image Understanding Workshop*, 1996.