2012 IEEE International Conference on
Multisensor Fusion and Integration for Intelligent Systems (MFI)
September 13-15, 2012. Hamburg, Germany

Bayesian 3D Independent Motion Segmentation with IMU-aided RBG-D Sensor

Jorge Lobo[1], João Filipe Ferreira[1], Pedro Trindade[2] and Jorge Dias[1,3]

Abstract— **In this paper we propose a two-tiered hierarchical Bayesian model to estimate the location of objects moving independently from the observer. Biological vision systems are very successful in motion segmentation, since they efficiently resort to flow analysis and accumulated prior knowledge of the 3D structure of the scene. Artificial perception systems may also build 3D structure maps and use optical flow to provide cues for ego- and independent motion segmentation. Using inertial and magnetic sensors and an image and depth sensor (RGB-D) we propose a method to obtain registered 3D maps, which are subsequently used in a probabilistic model (the bottom tier of the hierarchy) that performs background subtraction across several frames to provide a prior on moving objects. The egomotion of the RGB-D sensor is estimated starting with the angular pose obtained from the filtered accelerometers and magnetic data. The translation is derived from matched points across the images and corresponding 3D points in the rotation-compensated depth maps. A gyro-aided Lucas Kanade tracker is used to obtain matched points across the images. The tracked points can also used to refine the initial sensor based rotation estimation. Having determined the camera egomotion, the estimated optical flow assuming a static scene can be compared with the observed optical flow via a probabilistic model (the top tier of the hierarchy), using the results of the background subtraction process as a prior, in order to identify volumes with independent motion in the corresponding 3D point cloud. To deal with the computational load CUDA-based solutions on GPUs were used. Experimental results are presented showing the validity of the proposed approach.**

I. INTRODUCTION

Motion cues play an essential part in perception – they are ubiquitous in the process of making sense of the surrounding world, both for humans and for robots. However, motion perception has been long considered a difficult problem to tackle in artificial perception; although there has been a substantial amount of work in attempting to devise a solution by solely using vision, the challenges faced by the need to distinguish between optical flow caused by self-motion of the observer (i.e. egomotion) and by objects or agents moving independently from the observer are not at all trivial.

In biological vision systems both static and dynamic inertial cues provided by the vestibular system also play an important role in perception. In particular, they are deeply involved in the process of motion sensing, and are fused with vision in the early processing stages of image processing (e.g, the gravity vertical cue). As a result, artificial perception systems for robotic applications have since recently been taking advantage from low-cost inertial sensors for complementing vision systems [1].

On the other hand, an interesting hypothesis has been raised by studies in neuroscience such as presented in [2], which states that there are fast routes in the brain that are used to rapidly paint the rough overall 3D view of an observed scene, which is then fed back to lower levels of 2D perceptual processing as a prior. In fact, it is also posited by several authors that an accumulated prior knowledge of the 3D structure of the scene is retroinjected into the primary brain sites for flow analysis, thus modulating motion segmentation processing.

Besides the work described in [1] and references therein, recent work has been done in reexamining the Lucas-Kanade method for real-time independent motion detection [3].

In our approach we combine, in a probabilistic way, an inter-frame estimate of independent motion, based on the difference between observed optical flow and the estimated optical flow given the scene depth map and observer ego-motion, with a background subtraction method based on the repeated observation of the same scene, to have a more robust independent motion segmentation.

The next section presents our approach for estimating the observer egomotion and registering the observed 3D point clouds to a common frame of reference. In section 3 the two-tiered Bayesian hierarchical model for independent motion segmentation is presented, combining background subtraction with optical flow consistency. This is followed by some experimental results and concluding remarks.

II. ESTIMATING EGOMOTION AND REGISTERING 3D POINT CLOUDS OF THE IMU-AIDED RGB-D SENSOR

A. Estimating and Compensating for Egomotion

A moving RGB-D observer of a background static scene with some moving objects computes at each instant a dense depth map (or point cloud) corresponding to the captured image. The point clouds will change in time due to both the moving objects and the observer ego-motion. A first step to process the incoming data is to register the point clouds to a common fixed frame of reference $\{\mathcal{W}\}$, as shown on Figure 1.

*The research leading to these results has been partially supported by the HANDLE project, which has received funding from the European Community's 7th Framework Programme under grant agreement ICT 231640.

[1]Jorge Lobo, João Filipe Ferreira and Jorge Dias are with the Institute of Systems and Robotics (ISR) and the Department of Electrical and Computer Engineering, University of Coimbra, Portugal {jlobo, jfilipe, jorge}@isr.uc.pt

[2]Pedro Trindade is with the ISR - Institute of Systems and Robotics, University of Coimbra, Portugal pedrotrindade@isr.uc.pt

[3]Jorge Dias is also with the Robotics Institute, Khalifa University, Abu Dhabi, UAE.

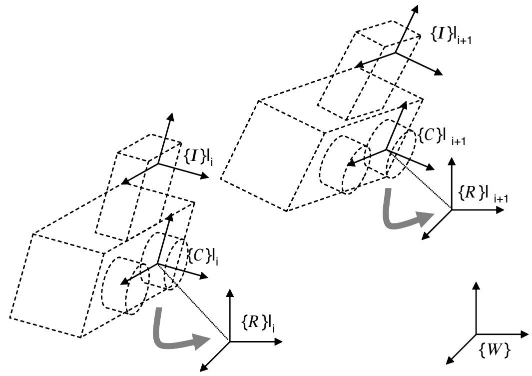

Fig. 1. Moving observer and world fixed frames of reference.

A set of 3D points ${}^{\mathcal{C}}\mathbb{P}|_i$ is therefore obtained at each frame, given in the camera frame of reference $\{\mathcal{C}\}|_i$. Each 3D point has RGB values corresponding to the intensity of the red, green and blue colour components, given by the colour pixel in the reference camera. Each point in the set retains both 3D position and colour component level, i.e.:

$$\boldsymbol{P}(x, y, z, r, g, b) \in {}^{\mathcal{C}}\mathbb{P}|_i \ . \tag{1}$$

The RGB-D sensor needs to be properly calibrated so that the correct correspondence is established between the image colour pixels and the 3D points in the point cloud.

1) Rotate to Local Vertical and Magnetic North: The inertial vertical reference alone could be used to rotate depth maps to a levelled frame of reference. However there remains a rotation about a vertical axis for which gravity provides no cues. The earth's magnetic field can be used to provide the missing bearing [4], however the magnetic sensing is sensitive to the nearby ferrous metals and electric currents. In fact, there is some overlap and complementarity between the two sensors, with different noise characteristics that can be exploited to provide a useful rotation update [5] [6].

The inertial and magnetic sensors, rigidly fixed to the depth camera rig, provide a stable camera rotation update ${}^{\mathcal{R}}\boldsymbol{R}_{\mathcal{C}}$ relative to the local gravity vertical and magnetic north camera frame of reference $\{\mathcal{R}\}|_i$.

Calibration of the rigid body rotation between $\{\mathcal{I}\}|_i$ and $\{\mathcal{C}\}|_i$ can be performed by having both sensors observing gravity, such as vertical vanishing points and sensed acceleration, as described [7].

The rotated camera frame of reference $\{\mathcal{R}\}|_i$ is time-dependent only due to the camera system translation, since rotation has been compensated for.

2) Translation from Image Tracked Features: The translation component can be obtained using a single fixed target tracked in the scene, or a set of tracked features to improve robustness. The image features must have the corresponding 3D point $\boldsymbol{P}_t$ in each depth map, so that translation can be estimated from

$$\Delta\vec{t} = \boldsymbol{P}_t|_{i+1} - \boldsymbol{P}_t|_i \tag{2}$$

with $\boldsymbol{P}_t|_{i+1} \in {}^{\mathcal{R}}\mathbb{P}|_{i+1}$ and $\boldsymbol{P}_t|_i \in {}^{\mathcal{R}}\mathbb{P}|_i$.

A set of sparse tracked natural 3D features can be used to improve robustness, but some assumptions have to be made in order to reject outliers that occur from tracking features of the moving objects. For this work we used a gyro-aided Luca Kanade tracker , running on a GPU using CUDA based code [8] [9]. The underlying assumption is made that the independent motion in the scene is not dominant, i.e., that the majority of the tracked features are from the observed static scene. This can later be improved by masking out regions where independent motion was observed and also take into account the ego motion of the observer, so that the tracked features can provide an estimate of the translation more reliably.

B. Occupancy Grid for 3D Point Cloud Registration

Registration of the acquired 3D point clouds was achieved by using an occupancy grid $\mathcal{Y}$ – a regular 3D Cartesian tesselation of cells (i.e. voxels), each indexed by C, coupled with an occupancy field associating each cell to a binary random variable O_C signalling the respective occupancy state.

Let $Z \equiv \cap_{i=1}^{N} Z_i$ represent the conjunction of the set of discretised readings corresponding to N points (x_i, y_i, z_i) composing the point cloud obtained by the range sensor, assumed to be *conditionally independent measurements*. The occupancy grid is to be updated by inferring $P(O_C|Z, M_C, Z_C)$ for each C, through the application of Bayes rule and marginalisation to the standard decomposition equation

$$P(O_C, D, Z, M_C, Z_C) = \\ P(D)P(O_C)P(Z_C|O_C, D)\prod_{i=1}^{N} P(M_C^i)P(Z_i|M_C^i, O_C, D), \tag{3}$$

where $M_C \equiv \cap_{i=1}^{N} M_C^i$ is the conjunction of N random variables M_C^i that signal if the corresponding Z_i falls within the limits of cell C, Z_C signals if there are *any* points within set Z falling within the limits of cell C, and finally D represents a binary random variable signalling either "detection" or "misdetection". The distributions involved in the decomposition are defined in the following lines.

The prior distribution $P([D = 0]) = P_{miss}$, $P([D = 1]) = 1 - P_{miss}$ introduces a meaningful error model that avoids deadlocks caused by 0 or 1 probabilities of occupancy, with P_{miss} being attributed an empirically chosen value; it also establishes the amount of inertia of the model with respect to changing the occupancy state of a cell after consecutive updates of the grid. The distribution $P(O_C)$ represents the prior on occupancy, taken from the posterior estimated in the previous time instant. Each distribution $P(M_C^i)$ represents a uniform (uninformative) prior.

The likelihood $P(Z_i|M_C^i, O_C, D)$ represents the direct sensor model of the generative formulation of the occupancy grid given by a delta Dirac distribution displaced to $Z_i = C$ if $M_C^i = 1$ and $D = 1$, or a uniform distribution $\mathcal{U}(Z_i)$

otherwise. Finally, the likelihood $P(Z_C|O_C, D)$ represents the probability of $O_C = 0$ implying that no measurement is falling within the limits of cell C; it is given by $P(Z_C|[O_C = 0], [D = 1]) = Z_C$, or a uniform distribution otherwise.

III. TWO-TIERED BAYESIAN HIERARCHICAL MODEL FOR INDEPENDENT MOTION SEGMENTATION

A. Bottom Tier – Bayesian Model for Background Subtraction

Background subtraction is performed by updating an inference grid similar to the occupancy grid described in section II-B, but, instead of occupancy, relating to the presence/absence of independent moving objects in cell C, represented by the binary random variable I_C. The rationale of background subtraction in this context is as follows: static objects will contribute with a steady influx of consistent readings registered in the occupancy grid, while moving objects will contribute with momentary, inconsistent readings. This will theoretically result in voxel cells associated with more certain states of occupancy corresponding to the static background, and any incoming reading inconsistent with these states will stand out as most probably having been caused by an independently moving object. The formal details of this process are presented next.

The independent motion grid of voxel cells is updated by inferring $P(I_C|Z, M_C, Z_C)$ for each C, through the application of Bayes rule and marginalisation to the decomposition equation

$$\begin{aligned} P(I_C, O_C^{-1}, D, Z, M_C, Z_C) = \\ P(D)P(O_C^{-1})P(I_C|O_C^{-1})P(Z_C|I_C, D) \\ \prod_{i=1}^{N} P(M_C^i)P(Z_i|M_C^i, I_C, D), \end{aligned} \tag{4}$$

where all variables (and respective distributions) are otherwise equivalent or analogous to the decomposition equation of the occupancy grid, excepting O_C^{-1}, which represents the occupancy of cell C in the *previous* inference step, and I_C.

The newly introduced distributions are defined as follows: $P(O_C^{-1})$ corresponds to the respective preceding posterior distribution of the occupancy grid); $P(I_C|O_C^{-1})$ is an inverse transition matrix, for which probability is maximal when $I_C \neq O_C^{-1}$ and minimal otherwise; and $P(Z_i|M_C^i, I_C, D)$ and $P(Z_C|I_C, D)$ have the same form as $P(Z_i|M_C^i, O_C, D)$ and $P(Z_C|O_C, D)$ for the occupancy grid, respectively, replacing O_C by I_C.

This means that the inference grid model works by labelling whatever object perceived by the range sensor that does not comply with the static background that has previously been mapped into the occupancy grid (i.e. $I_C \neq O_C^{-1}$) *as an independently moving object*.

B. Top Tier – Bayesian Model for Optical Flow Consistency-Based Segmentation

Optical flow is the apparent motion of brightness patterns in the image. Generally, optical flow corresponds to the projected motion field, but not always. Shading, changing lighting and some texture patterns might induce an optical field different from the motion field. However since what can be observed is the optical field, the assumption is made that optical flow field provides a good estimate for the true projected motion field.

Optical flow computation can be made in a *dense* way, by estimating motion vectors for every image pixel, or *feature based*, estimating motion parameters only for matched features.

The camera provides colour intensity images $I(u, v)|_i$ where u and v are pixel coordinates, and i the frame time index. Each point has an RGB value and a corresponding intensity gray level. Representing the 2D velocity of an image pixel $\boldsymbol{u} = (u, v)^\intercal$ as $\frac{d\boldsymbol{u}}{dt}$, the brightness constancy constraint says that the projection of a world point has a constant intensity over a short interval of time, i.e., assuming that the pixel intensity or brightness is constant during dt, we have

$$I(u + \frac{du}{dt}dt, v + \frac{dv}{dt}dt)|_{t+dt} = I(u, v)|_t \tag{5}$$

If the brightness changes smoothly with u, v and t, we can expand the left-hand-side by a Taylor series and reject the higher order terms to obtain

$$\nabla \boldsymbol{I} \cdot \frac{d\boldsymbol{u}}{dt} + \frac{\partial I}{\partial t}dt = 0 \tag{6}$$

where ∇I is the image gradient at pixel $\boldsymbol{u}$. These spatial and time derivatives can be estimated using a convolution kernel on the image frames.

But for each pixel we only have one constraint equation, and two unknowns. Only the *normal flow* can be determined, i.e., the flow along the direction of image gradient. The flow on the tangent direction of an isointensity contour cannot be estimated. This is the so called *aperture problem*. Therefore, to determine optical flow uniquely additional constraints are needed.

The problem is that a single pixel cannot be tracked, unless it has a distinctive brightness with respect to all of its neighbours. If a local window of pixels is used, a local constraint can be added, i.e., single pixels will not be tracked, but windows of pixels instead.

Barron *et al.* [10] present a quantitative evaluation of optical flow techniques, including the Lucas-Kanade method, that uses local consistency to overcome the aperture problem [11]. The assumption is made that a constant model can be used to describe the optical flow in a small window.

When the camera is moving and observing a static scene with some moving objects, some optical flow will be consistent with the camera ego-motion observing the static scene, other might be moving objects. Since we have the 3D scene dense depth map, and we reconstruct camera motion, we can

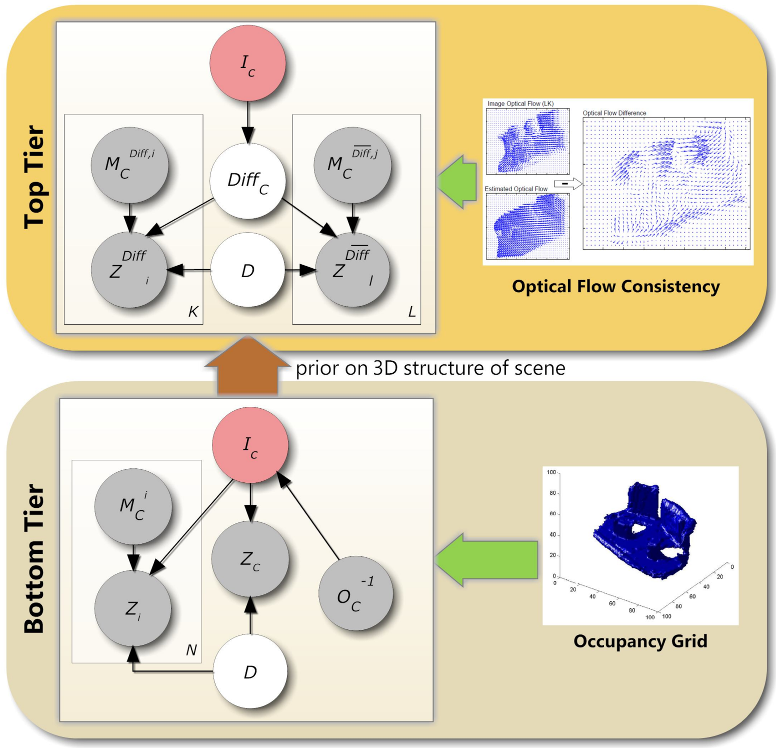

Fig. 2. Full hierarchical framework for independent motion segmentation. Bayesian networks using plate notation [12] corresponding to each of the hierarchy tiers are presented, with searched variables in red, hidden/unwanted variables to marginalise with no fill and measured variables in grey.

compute the expected projected optical flow in the image from the 3D data.

In the perspective camera model, the relationship between a 3D world point $\boldsymbol{x} = (X, Y, Z)^{\mathsf{T}}$ and its projection $\boldsymbol{u} = (u, v)^{\mathsf{T}}$ in the 2D image plane is given by

$$u = \frac{\boldsymbol{P}_1 \, (x, y, z, 1)^{\mathsf{T}}}{\boldsymbol{P}_3 \, (x, y, z, 1)^{\mathsf{T}}} \qquad v = \frac{\boldsymbol{P}_2 \, (x, y, z, 1)^{\mathsf{T}}}{\boldsymbol{P}_3 \, (x, y, z, 1)^{\mathsf{T}}} \tag{7}$$

where matrix $\boldsymbol{P}_j$ is the jth row of the camera projection matrix $\boldsymbol{P}$.

When the camera moves, the relative motion of the 3D point $\frac{d\boldsymbol{x}}{dt}$ will induce a projected optical flow given by

$$\frac{d\boldsymbol{u}_i}{dt} = \frac{\delta \boldsymbol{u}_i}{\delta \boldsymbol{x}} \frac{d\boldsymbol{x}}{dt} \tag{8}$$

where $\frac{\delta \boldsymbol{u}_i}{\delta \boldsymbol{x}}$ is the 2×3 Jacobian matrix that represents the differential relationship between $\boldsymbol{x}$ and $\boldsymbol{u}_i$, which can be obtained by differentiating (7).

Image areas where the computed flow is inconsistent with the expected one indicate moving objects, and the corresponding voxels in the cell grid can be segmented.

The difference image between the estimated and the measured optical flow is then thresholded and binarised. Consequently, two mutually exclusive sets of random variables of the same form as Z can be defined, Z^{Diff} and $Z^{\overline{Diff}}$, by classifying points from the cloud yielded by the range sensor as either corresponding to a *non-consistent pixel* or to a *consistent pixel* with corresponding variables analogous to M_C, M_C^{Diff} and $M^{\overline{Diff}_C}$, respectively.

Using these random variables, the top-level inference grid is updated by inferring $P(I_C | Z^{Diff}, Z^{\overline{Diff}}, M_C^{Diff}, M_C^{\overline{Diff}})$ for each voxel cell C, through the application of Bayes rule and marginalisation to the decomposition equation

$$\begin{aligned} & P(I_C, Z^{Diff}, Z^{\overline{Diff}}, M_C^{Diff}, M_C^{\overline{Diff}}, D, Diff_C) = \\ & \quad P(D) P(I_C) P(Diff_C | I_C) \\ & \quad \prod_{i=1}^{K} P(M_C^{Diff,i}) P(Z_i^{Diff} | Diff_C, M_C^{Diff,i}, D) \\ & \quad \prod_{j=1}^{L} P(M_C^{\overline{Diff},j}) P(Z_j^{\overline{Diff}} | Diff_C, M_C^{\overline{Diff},j}, D), \end{aligned} \tag{9}$$

where the remaining random variables have the same meaning as before, with the exception of $Diff_C$, a hidden binary variable which signals if a cell C is labelled as being occupied by an independently moving object, *if considering consistency-based segmentation*.

Since it is expected that consistency-based segmentation and background subtraction segmentation yield the same results, the distribution $P(Diff_C | I_C)$ is simply a transition matrix for which probability is maximal when $Diff_C = I_C$ and minimal otherwise. The distribution $P(I_C)$ provides *the link between the two tiers of the hierarchy*, and is given by the result of inference on the lower level, $P(I_C | Z, M_C, Z_C)$. It models the *accumulated prior knowledge of the 3D structure of the scene*, thus representing an analogous process to what is believed to happen in the human brain, as described in the introductory section.

Finally, $P(D)$ and $P(M_C^{Diff,i})$, $P(M_C^{\overline{Diff},j})$ and $P(Z_i^{Diff}|Diff_C, M_C^{Diff,i}, D)$ follow analogous definitions to the corresponding distributions in previous models, while $P(Z_j^{\overline{Diff}}|Diff_C, M_C^{\overline{Diff},j}, D)$ is given by a delta Dirac distribution displaced to $Z_j^{\overline{Diff}} = C$ for $Diff_C = 0$, $M_C^{\overline{Diff},j} = 1$ and $D = 1$, or a uniform distribution $\mathcal{U}(Z_i)$ otherwise.

The full hierarchical framework is presented on Fig. 2. The posterior of the top tier of the hierarchy only needs to be inferred up to a proportion of the product of the nonuniform priors and likelihoods, to then apply a maximum a posteriori (MAP) decision rule in order to estimate the segmented independent motion. Conversely, the posterior distributions of the occupancy grid and the bottom tier of the hierarchy should be exactly inferred; however, the respective models have been designed so that inference can be easily and efficiently performed using closed-form solutions.

IV. RESULTS

Using a MS Kinect as the RGB-D sensor, and attaching a Xsens MTix IMU sensor, that has both inertial and magnetic sensors, we were able to acquire datasets with images, corresponding 3D point clouds or depth maps, and rotation update. Fig. 3 shows the setup used, where optotracker markers where added to provide ego-motion ground truth, to be used later for benchmarking and refining the implemented method.

Figure 4 shows preliminary results where there is a moving object swinging by a static background scene, and the observer is moving while surveying the scene. On the left we can see the prior for background subtraction, that corresponds to the 3D voxels that were repeatedly observed and probably belong to the observed static scene. In the centre we have the unfiltered output of independent motion segmentation based on background segmentation from the bottom tier. On the right we can see the final filtered top tier result. Figure 4 only shows a single frame; the full sequence is best observed on the accompanying video. In the video we can see that the method works to some extent in this controlled scene. Further testing is required to evaluate the behaviour of the method under more challenging situations, where the underlying assumptions are not always fully met.

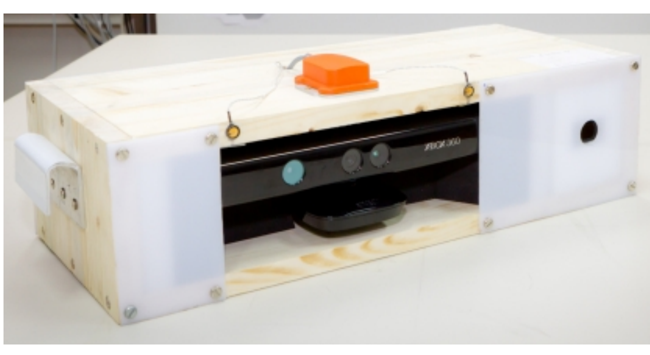

Fig. 3. Experimental setup with RGB-D (MS Kinect) and IMU (Xsens MTix) sensors.

V. CONCLUSIONS

In this paper we proposed a two-tiered hierarchical Bayesian model to estimate the location of objects moving independently from the observer. Using a RGB-D sensor with an attached IMU, we were able to have a rotation update from the filtered accelerometer and magnetic data that, combined with tracked features on the image sequence, provided an estimate for the egomotion of the sensors. This allowed the estimation of the optical flow assuming the observed scene was static, and miss-matches with the observed flow provided indication of independent motion. Using the temporal sequence to construct a prior on the scene's static background, the implemented probabilistic model combines this with the optical flow miss-match to find voxels with independent motion.

It is clear that the probabilistic fusion of background subtraction prior and optical flow consistency works to some extent, outperforming the isolated approaches. However further work is needed to deal with edge effects and remaining noise. The main source of both problems is the fact that we are modelling the absence of a sensor reading signalling a 3D point from the depth map falling within a cell C with a likelihood that decays the belief of occupancy of that cell. Although this tends to remove the effect of erroneous readings and reinforce correct measurements, static objects detected previously which subsequently fall outside the field of view (i.e., due to sensor egomotion) will eventually be "forgot" by the model.

However, the depth sensor used in this work functions, in fact, as an array of linear depth sensors; these sensors project rays that traverse empty space until there is a reflection on an object surface. This means that the RGB-D sensor not only provides readings relating to occupancy, but it also provides evidence of *empty space* between the sensor and the detected surface, which could be used to replace the "forgetfulness" likelihood approach. In future work, we propose to devise a more sophisticated model of the depth sensor so as to take advantage of this property.

REFERENCES

[1] P. Corke, J. Lobo, and J. Dias, "An introduction to inertial and visual sensing," *The International Journal of Robotics Research (IJRR) Special Issue from the 2nd Workshop on Integration of Vision and Inertial Sensors.*, vol. 26, no. 6, pp. 519–535, June 2007. [Online]. Available: http://ijr.sagepub.com/cgi/content/abstract/26/6/519

[2] J. Bullier, "Integrated model of visual processing," *Brain Research Reviews*, vol. 36, no. 2–3, pp. 96–107, Oct. 2001. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165017301000856

[3] C. Ciliberto, U. Pattacini, L. Natale, F. Nori, and G. Metta, "Reexamining Lucas-Kanade method for real-time independent motion detection: Application to the iCub humanoid robot," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, Sept. 2011, pp. 4154 –4160.

[4] M. J. Caruso, T. Bratland, C. H. Smith, and R. Schneider, "A New Perspective on Magnetic Field Sensing," Honeywell, Inc, Tech. Rep., 1998. [Online]. Available: http://www.ssec.honeywell.com/magnetic/datasheets/new_pers.pdf

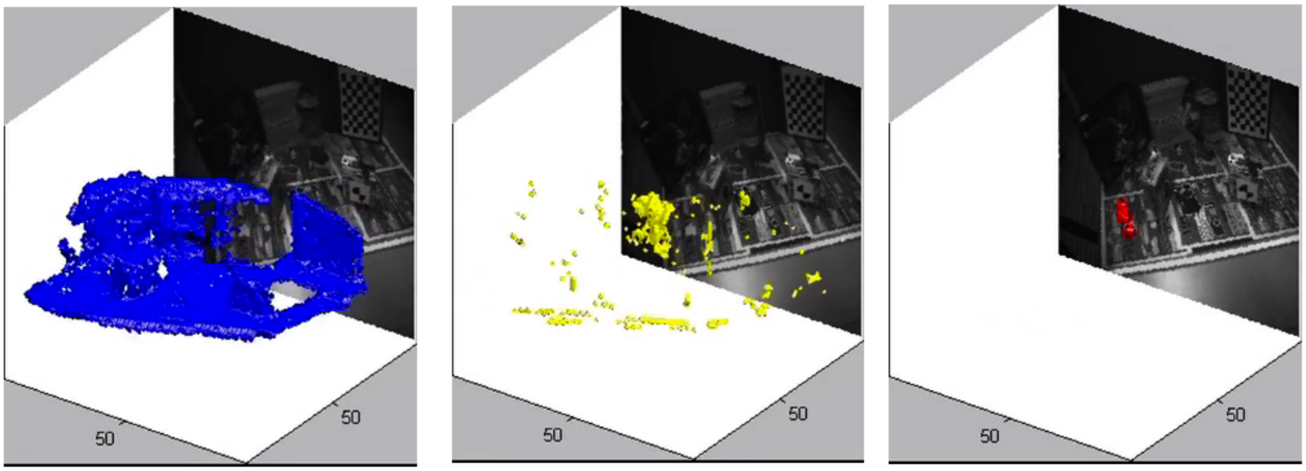

Fig. 4. Results showing background subtraction prior (blue on the left), the unfiltered background subtraction bottom tier result (yellow on the centre) and the filtered final top tier result (red on the right).

[5] D. Roetenberg, H. Luinge, and P. Veltink, "Inertial and magnetic sensing of human movement near ferromagnetic materials," in *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality*, October 2003, pp. 268–269.

[6] D. Roetenberg, H. Luinge, C. Baten, and P. Veltink, "Compensation of magnetic disturbances improves inertial and magnetic sensing of human body segment orientation," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on [see also IEEE Trans. on Rehabilitation Engineering]*, vol. 13, no. 3, pp. 395–405, September 2005. [Online]. Available: http://ieeexplore.ieee.org/search/wrapper.jsp?arnumber=1506825

[7] J. Lobo and J. Dias, "Relative pose calibration between visual and inertial sensors," *The International Journal of Robotics Research (IJRR) Special Issue from the 2nd Workshop on Integration of Vision and Inertial Sensors.*, vol. 26, pp. 561–577, 2007.

[8] M. Hwangbo, J.-S. Kim, and T. Kanade, "Inertial-aided klt feature tracking for a moving camera," in *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, 2009, pp. 1909 –1916. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5354093

[9] J.-S. Kim, M. Hwangbo, and T. Kanade, "Realtime affine-photometric klt feature tracker on gpu in cuda framework," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, 272009-oct.4 2009, pp. 886 –893. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5457608

[10] J. Barron, D. Fleet, and S. Beauchemin, "Performance of Optical Flow Techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, February 1994.

[11] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision ," in *Proceedings of Imaging Understanding Workshop*, 1981, pp. 674–679.

[12] W. L. Buntine, "Operations for learning with graphical models," *J. Artif. Int. Res.*, vol. 2, no. 1, p. 159–225, Dec. 1994. [Online]. Available: http://dx.doi.org/10.1613/jair.62