Tracking of multiple objects in unknown background using Bayesian estimation in 3D space

Yige Zhao, Xiao Xiao, Myungjin Cho, and Bahram Javidi*

Electrical and Computer Engineering Department, University of Connecticut, Storrs, Connecticut 06269-2157, USA
**Corresponding author: bahram.javidi@uconn.edu*

Received January 24, 2011; revised July 13, 2011; accepted July 14, 2011;
posted August 2, 2011 (Doc. ID 140610); published August 29, 2011

We present a three-dimensional (3D) object tracking method based on a Bayesian framework for tracking multiple, occluded objects in a complex scene. The 3D passive capture of scene data is based on integral imaging. The statistical characteristics of the objects versus the background are exploited to analyze each frame. The algorithm can work with objects with unknown position, rotation, scale, and illumination. Posterior probabilities of the reconstructed scene background and the 3D objects are calculated by defining their pixel intensities as Gaussian and gamma distributions, respectively, and by assuming appropriate prior distributions for estimated parameters. Multiobject tracking is achieved by maximizing the geodesic distance between the log-posteriors of the background and the objects. Experimental results are presented.

OCIS codes: 110.6880, 280.4991, 150.6910.

1. INTRODUCTION

Three-dimensional (3D) tracking of multiple objects in a scene is of interest in many areas, including surveillance, robotics, and security. In some cases, objects of interest may be partially occluded, making tracking with two-dimensional (2D) images difficult due to the superposition of occlusion noise and object details. Tracking with 3D imaging offers advantages over 2D imaging systems because of its robustness to object occlusion and the possibility to track multiple objects moving in all 3D coordinates, including range estimation. Also, tracking may need to be robust to variations in object or background features, such as variations in object orientation and scene illumination.

There have been numerous approaches to address detection, recognition, and tracking problems using 3D integral imaging [1–7] or multiperspective imaging [8], or other approaches. One possible solution is contour-based object tracking [9–13]. Detection of the object is required for these methods, and then tracking is conducted by moving the previous contour toward the current boundaries. In light of the fact that the active contour method [13] evaluates the changes of local intensities along the boundary; it is limited to small displacements. On the other hand, region-based methods [10,11] exploit the information of both the object and the background for more robust and flexible performance.

In this paper, we present tracking of multiple occluded 3D objects using a region tracking method based on statistical Bayesian formulation and 3D integral imaging used for passive sensing and computational 3D scene reconstruction. It is assumed that the background is stationary for each frame. We also assume that the reconstructed pixel intensities of both background and multiple objects are independent identically distributed (IID), and they follow Gaussian and gamma distributions based on their grayscale images, respectively. Within the Bayesian framework, posterior probabilities of background and objects are calculated by assuming the appropriate prior distributions for estimated parameters. At each incoming frame, the 3D scene is reconstructed. Then, the objects are located in 2D slices of the 3D reconstructed scene by maximizing the geodesic distance [14] between the log-posteriors of the reconstructed background and objects to be tracked. Then, each object is tracked individually in 3D space by maximizing the above distance across all the 2D reconstructed planes.

In Section 2, we briefly describe the concepts of 3D passive image sensing and visualization. Our statistical Bayesian tracking algorithm is presented in Section 3. The experimental results are demonstrated in Section 4, followed by the summary and conclusion.

2. SYNTHETIC APERTURE INTEGRAL IMAGING AND COMPUTATIONAL RECONSTRUCTION

As illustrated in Fig. 1(a), a camera array is used to acquire the elemental images from slightly different perspectives with respect to the scene. For computational reconstruction, each elemental image is projected through an associated virtual pinhole array to the desired reconstruction plane, and it is superimposed with other projected elemental images [6]. The 3D scene can be computationally reconstructed plane by plane using this method, as depicted in Fig. 1(b). Each elemental image is projected and superimposed by the magnification $M = d/f$, where d is the distance from the image sensor to the 3D object and f is the focal length of the image sensor, respectively. This enables visualization of the partially occluded objects, because only the reconstruction plane with the object of interest is in focus, while the occlusion and background are out of focus.

3. TRACKING WITH THE BAYESIAN ALGORITHM

Object segmentation is applied on the reconstructed images for tracking. Reconstructed images are divided into

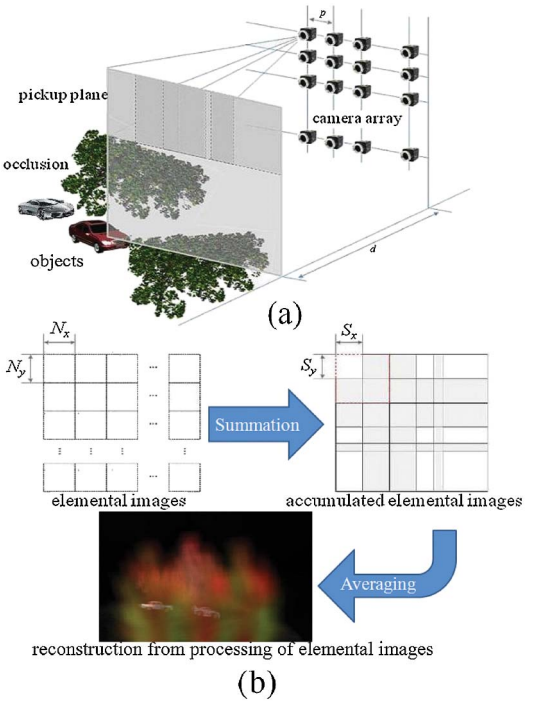

Fig. 1. (Color online) 3D integral imaging sensing and reconstruction: (a) scene capture process and (b) 3D reconstruction of the scene in Fig. 2.

the background region (Ω_b) and the object region (Ω_o) [15–17]. Our goal is to find the object region Ω_o (in a statistical sense), matching the object support. We assume that the background and the objects are statistically independent and that the background is stationary for each frame. The objects' pixel intensities are usually correlated. However, for simplicity, we assume that the pixel intensities of both the reconstructed background and the reconstructed objects are unknown, independent, and follow Gaussian and gamma distributions, respectively. We will present experimental results in Section 4 illustrating the tracking performance under these assumptions. In the following derivations, for simplicity, one-dimensional notations are used for the signals as $\mathbf{s} = \{s_i | i \in [1, N]\}$, where N is the total number of pixels. Let $\mathbf{w} = \{w_i | i \in [1, N]\}$ be a binary window that defines a support for objects, such that $w_i = 1$ for object pixels (denoted as $\mathbf{o}$), and $w_i = 0$ for background pixels (denoted as $\mathbf{b}$). The purpose of segmentation is to estimate the window function $\mathbf{w}$ for objects of interest in the reconstructed scene. Thus, each point on the reconstruction can be modeled as a spatially disjoint combination of object and background as

$$s_i = o_i w_i + b_i(1 - w_i). \tag{1}$$

Several optimal criterion laws [18] have been derived for situations that the statistical properties of object and background are from the exponential family (i.e., gamma or Gaussian). In 3D integral imaging reconstruction of the scene, the optical rays generated by elemental images are superimposed. Thus, the background region of the reconstructed images tends to be Gaussian distributed by applying the central limit theorem. The statistical behavior of various objects may be different. Therefore, a gamma distribution is chosen as a robust statistical distribution to capture the object pixel distributions. By adjusting its parameters, we can approximate various distributions. Also, the object statistics at different times (frames) or at different object poses or orientations may vary. The gamma distribution parameters can be estimated to capture such variations of the object.

A. Background Region Statistics

By assuming a Gaussian distribution for the background region, one can write the probability density function (PDF) as follows:

$$f_b(\mathbf{s}) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mathbf{s}-\mu)^2}{2\sigma^2}}, \tag{2}$$

where μ and σ^2 are the mean and the variance for the background region, respectively.

We estimate the unknown variables μ and σ^2 by maximizing the conditional probability $\mathbf{P}_b(\mu, \sigma^2 | \mathbf{w}, \mathbf{s})$:

$$(\hat{\mu}, \hat{\sigma}^2) = \arg\max_{(\mu,\sigma^2)} P_b(\mu, \sigma^2 | \mathbf{w}, \mathbf{s}), \tag{3}$$

where $(\hat{\mu}, \hat{\sigma}^2)$ is the estimate of (μ, σ^2) in the maximum *a posteriori* (MAP) sense. According to Bayes's rule [19], the conditional probability can be rewritten as

$$\begin{aligned} P_b(\mu, \sigma^2 | \mathbf{w}, \mathbf{s}) &= \frac{P_b(\mathbf{s} | \mathbf{w}, \mu, \sigma^2) P_b(\mu, \sigma^2)}{P_b(\mathbf{s})} \\ &= \frac{P_b(\mathbf{s} | \mathbf{w}, \mu, \sigma^2) P_b(\mu) P_b(\sigma^2 | \mu)}{P_b(\mathbf{s})}. \end{aligned} \tag{4}$$

We assume that $P_b(\mu)$ is uniformly distributed and $P_b(\sigma^2 | \mu) \propto 1/\sigma^2$. Taking Eq. (1) into account, we write the likelihood $P_b(\mathbf{s} | \mathbf{w}, \mu, \sigma^2)$ as follows:

$$P_b(\mathbf{s} | \mathbf{w}, \mu, \sigma^2) = \prod_{i=1}^{N} \left[\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(s_i-\mu)^2}{2\sigma^2}} \right]^{(1-w_i)}. \tag{5}$$

In order to derive the MAP of unknown parameters, one has to take partial derivatives of the log-posterior function with respect to μ and σ^2, and set each to zero as follows:

$$\begin{aligned} &\frac{\partial \log P_b}{\partial \mu} = 0 \Rightarrow \hat{\mu} | \mathbf{s}, \mathbf{w} = \frac{1}{N_{(1-\mathbf{w})}} \sum_{i=1}^{N} s_i (1 - w_i), \\ &\frac{\partial \log P_b}{\partial \sigma^2} = 0 \Rightarrow \hat{\sigma}^2 | \hat{\mu}, \mathbf{s}, \mathbf{w} = \frac{1}{N_{(1-\mathbf{w})} + 2} \sum_{i=1}^{N} (s_i - \hat{\mu})^2 (1 - w_i), \end{aligned} \tag{6}$$

where $N_{(\cdot)}$ is an operator providing the number of ones in its operand and N is the total number of input image pixels.

B. Object Region Statistics

By assuming an IID gamma distribution for the pixels of the multiple 3D objects to be tracked, one can write the PDF of the object j as follows:

$$f_{oj}(\mathbf{s}) = \frac{\beta^\alpha}{\Gamma(\alpha)} \mathbf{s}^{\alpha-1} e^{-\beta \mathbf{s}}, \tag{7}$$

where j denotes the index of objects to be tracked, $\Gamma(\cdot)$ is the gamma function, and $\alpha > 0$, $\beta > 0$.

We assume that the shape parameter α is known, and that the rate parameter β has known gamma prior distribution, $\pi(\beta) \equiv \text{gamma}(\alpha_0, \beta_0)$ whose α_0 and β_0 are known. (α_0, β_0) are selected based on the types of objects and scenes used in the experiments. One can derive the *posterior* distribution of each object region:

$$\begin{aligned} P_{oj}(\beta|\alpha, \mathbf{w}, \mathbf{s}) &\propto P_{oj}(\mathbf{s}|\mathbf{w}, \alpha, \beta)\pi(\alpha_0, \beta_0) \\ &= \prod_{i=1}^{N}\left[\frac{\beta^{\alpha}}{\Gamma(\alpha)} s_i^{\alpha-1} e^{-\beta s_i}\right]^{w_i} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \beta^{\alpha_0-1} e^{-\beta_0 \beta} \\ &= \frac{\beta^{\alpha N_{(\mathbf{w})}}}{\Gamma(\alpha)^{N_{(\mathbf{w})}}} \left(\prod_{i=1}^{N} s_i^{w_i}\right)^{\alpha-1} e^{-\beta \sum_{i=1}^{N} s_i w_i} \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \beta^{\alpha_0-1} e^{-\beta_0 \beta} \\ &\propto \beta^{(\alpha N_{(\mathbf{w})}+\alpha_0)-1} e^{-\beta\left(\sum_{i=1}^{N} s_i w_i + \beta_0\right)} \\ &\sim \text{gamma}\left(\alpha N_{(\mathbf{w})} + \alpha_0, \sum_{i=1}^{N} s_i w_i + \beta_0\right). \end{aligned} \tag{8}$$

The posterior distribution of the object region is also gamma distributed; however, with different parameters. The Bayes' estimator of β under the squared error loss is achieved as the posterior mean [19]:

$$\hat{\beta}|\alpha, \mathbf{w}, \mathbf{s}, \alpha_0, \beta_0 = \frac{\alpha N_{(\mathbf{w})} + \alpha_0}{\sum_{i=1}^{N} s_i w_i + \beta_0}. \tag{9}$$

C. 3D Tracking with the Bayesian Algorithm

The tracking of objects can be modeled as an estimation problem. Our objective is to estimate the object subspace in a 3D stack of reconstructed planes obtained by integral imaging. Assume that the initial positions of our objects are located in some regions with unknown reconstructed planes in the 3D space. The objects are tracked individually, because they may be located at different depths. For the first frame, starting with an arbitrary reconstructed plane p between the occlusion and the background, we are first seeking to locate the objects individually, which is analogous to maximizing the geodesic distance [14] of the object j and the background:

$$\begin{aligned} &\varepsilon_{pj}(\mathbf{s}, \mathbf{w}) \\ &= \sqrt{E\left[\log\left(\frac{P_{oj}(\mathbf{s}|\mathbf{w}, \alpha, \hat{\beta})}{P_b(\mathbf{s}|\mathbf{w}, \hat{\mu}, \hat{\sigma}^2)}\right)^2\right] - \left\{E\left[\log\left(\frac{P_{oj}(\mathbf{s}|\mathbf{w}, \alpha, \hat{\beta})}{P_b(\mathbf{s}|\mathbf{w}, \hat{\mu}, \hat{\sigma}^2)}\right)\right]\right\}^2}, \end{aligned} \tag{10}$$

where p is the reconstructed plane and j is the index of objects to be tracked. Thus for an object j at the reconstructed plane p, $\mathbf{w}$ is the optimal binary window and the optimal segmentation.

Equation (10) can be interpreted as the snake energy [13]. The maximization is done by applying a level-set method [20]. Let $\Gamma(t)$ be the object surface at time t, the embedding level-set is defined as $\varphi(\mathbf{q}, t)$, where $\mathbf{q}$ denotes a point in the level-set, such that $\varphi(\mathbf{q}, t) < 0$ represents the object region, and $\varphi(\mathbf{q}, t) > 0$ represents the background region. Our object surface can be explicitly written as $\Gamma(t = t_0) = \{\mathbf{q}|\varphi(\mathbf{q}, t = t_0) = 0\}$. It can be shown [10,20] that if each point propagates to the interface $(\vec{n} = \nabla_\varphi/|\nabla_\varphi|)$, then the evolution of $\Gamma(t)$ can be modeled as a discrete space–time partial differential equation, $\varphi(\mathbf{q}, t+1) = \varphi(\mathbf{q}, t) + F(\mathbf{q})\|\vec{\nabla}_\varphi\|$, where $F(\mathbf{q})$ represent the speed function.

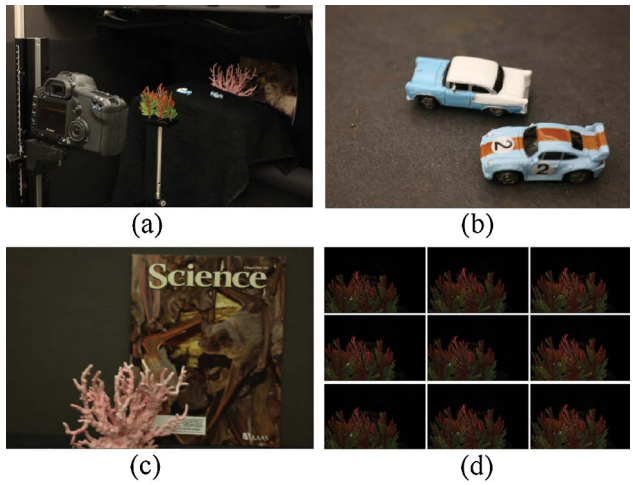

Fig. 2. (Color online) Experimental setup and objects with unknown occlusion and background used in the scene: (a) experimental setup, (b) objects to be tracked (two cars), (c) background, and (d) elemental images.

Thus the maximization problem is analogous to computing the derivatives of $\varepsilon_{pj}(\mathbf{s}, \mathbf{w})$ with respect to $\mathbf{s}$. The corresponding Euler–Lagrange equation result is $\partial\varepsilon_{pj}(\mathbf{s}, \mathbf{w})/\partial\mathbf{s} = (\varepsilon_{pj}(\mathbf{s}, \mathbf{w}))\vec{n}$, where $\vec{n}$ is the outward normal to the object surface. By following Ref. [10], the speed function can be rewritten as

$$F(\mathbf{q}) = \nabla_\varphi \varepsilon_{pj}(\mathbf{s}, \mathbf{w}) + \text{div}\left(\frac{\nabla(\varphi(\mathbf{q}, t))}{|\nabla(\varphi(\mathbf{q}, t))|}\right). \tag{11}$$

Then each object is tracked individually in 3D space by maximizing the distance in Eq. (10) across all the reconstructed planes of interest:

$$\varepsilon_j(\mathbf{s}, \hat{\mathbf{w}}, \hat{p}) = \arg\max_{(p,\mathbf{w})} \varepsilon_{pj}(\mathbf{s}, \mathbf{w}). \tag{12}$$

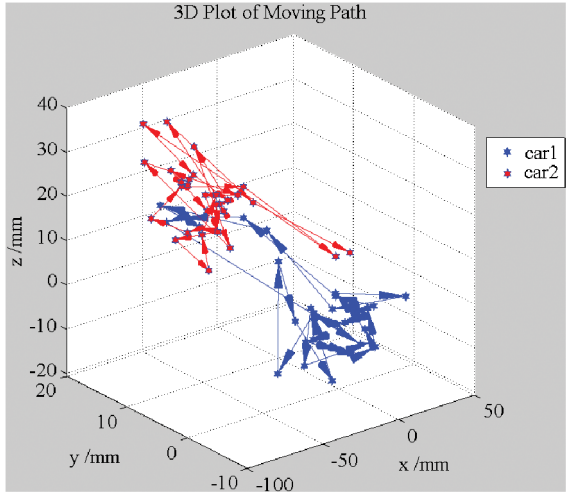

Fig. 3. (Color online) 3D plot of objects positions to be tracked (3D movements).

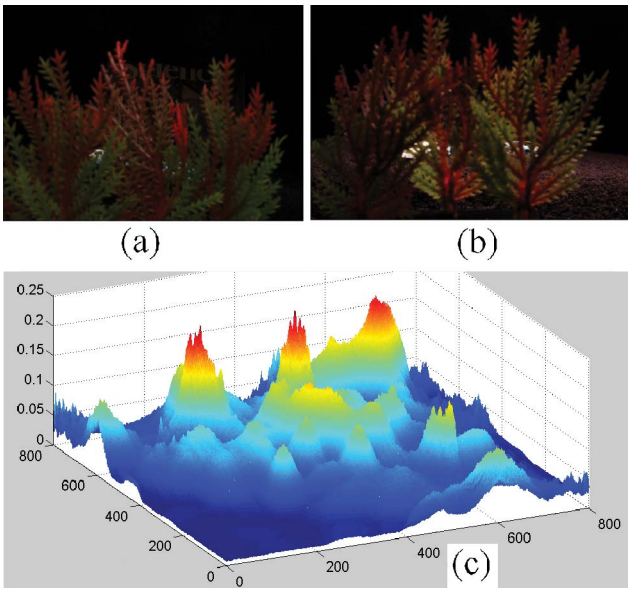

Fig. 4. (Color online) 2D tracking results of optimal object tracking algorithm presented in Ref. [17] with objects rotated and illumination changed for scenes in (a) and (b): (a) two occluded objects in frame two, (b) two occluded objects in frame three, and (c) tracking results of frame three.

4. EXPERIMENTAL RESULTS

For optical experiments, two cars with unknown position, rotation, illumination, and the presence of unknown occlusion and background are used as objects to be tracked [see Fig. 2(a)]. Objects are shown in Fig. 2(b), and the background is shown in Fig. 2(c). Elemental (multiview) images for this

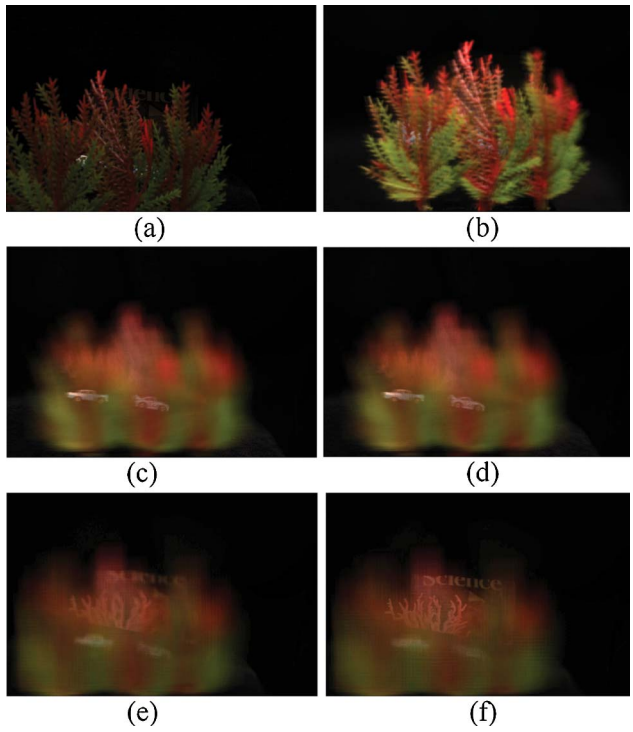

Fig. 5. (Color online) 3D reconstruction with occluded objects (a) occluded objects (cars behind tree branches) are located at $Z = 380$ mm and $Z = 410$ mm from the sensor, respectively. (b)–(f) 3D reconstructed image sequences at $Z = 190$, $Z = 380$, $Z = 410$, $Z = 570$, and $Z = 690$ mm, respectively. (b) Reconstructed plane of occlusion (tree branches).

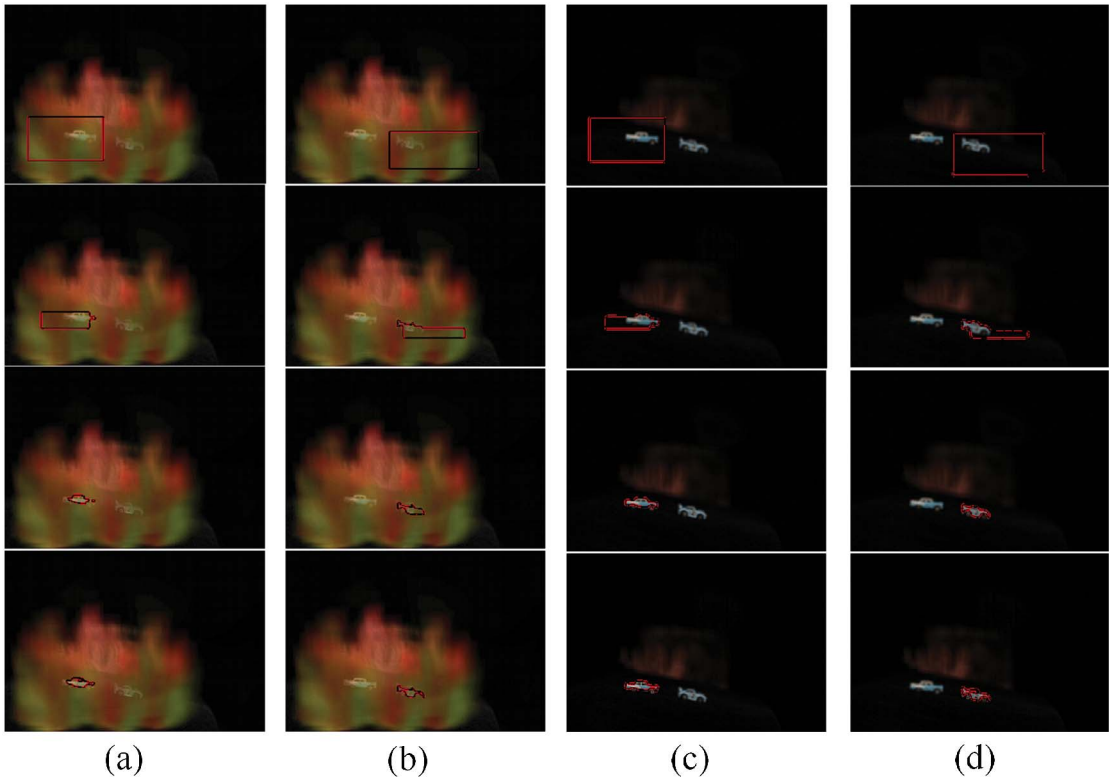

Fig. 6. (Color online) 3D tracking results for the first frame: (a) car 1 with occlusion (Media 1), (b) car 2 with occlusion (Media 2), (c) car 1 without occlusion (Media 3), and (d) car 2 without occlusion (Media 4)

scene are captured as illustrated in Fig. 1. Each elemental image has 2784×1856 pixels. Sample elemental images with various perspectives are shown in Fig. 2(d). A camera lens with $f = 50\,\text{mm}$ is used. Our camera is located at distance $Z = 0\,\text{mm}$; the occlusion is located at $Z = 190\,\text{mm}$; the two cars to be tracked are initially located at $Z = 380\,\text{mm}$ and $Z = 410\,\text{mm}$, respectively, and the background is located at $Z = 690\,\text{mm}$. The two objects (cars) are moved randomly, and 60 positions for both objects are recorded. The 3D movements of the objects are shown in Fig. 3.

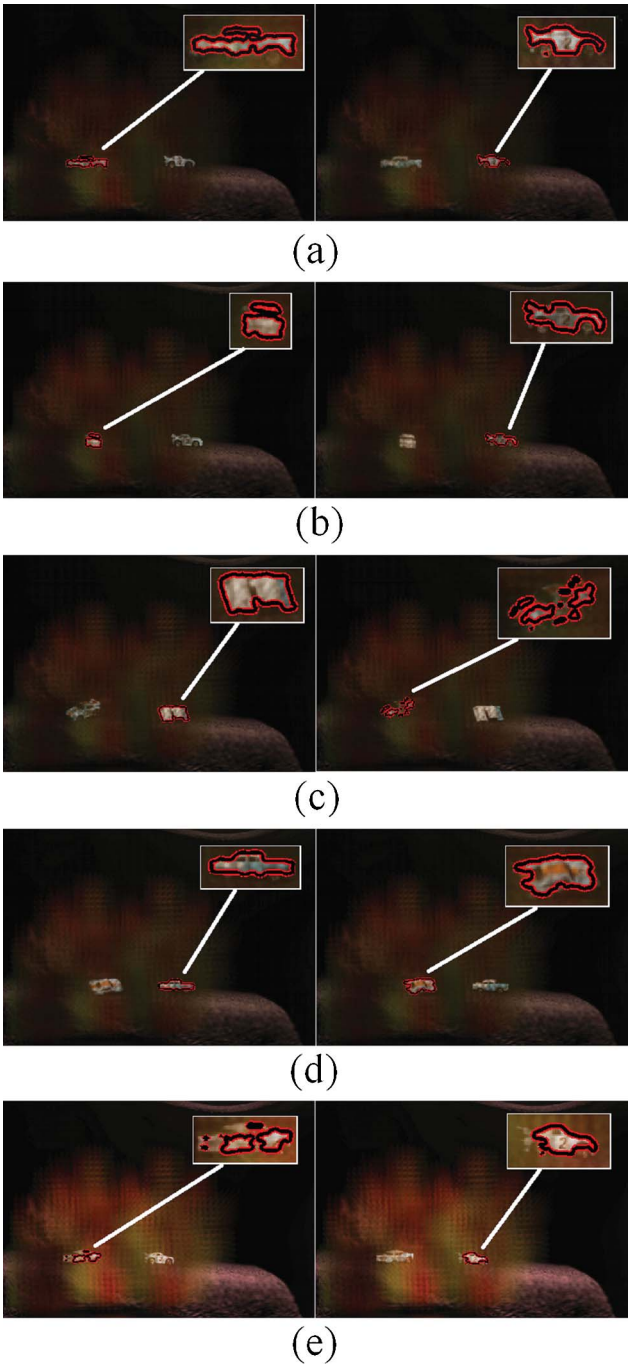

Fig. 7. (Color online) 3D tracking results of moving cars (Media 5 and Media 6) in unknown background (see Fig. 2). The objects' movements are shown in Fig. 3. (a) Frame 2, (b) frame 9 (illumination reduced to one-half and with car 1 rotated), (c) frame 14 (with both cars rotated), (d) frame 17 (with car 2 rotated), and (e) frame 27 (scene illumination doubled and with car 1 rotated).

Our background region is modeled as Gaussian distribution with mean μ and variance σ^2, whose values can be estimated from Eq. (6). Those estimates vary and depend on the scene and on the illumination conditions. For the object region, α may be different from object to object, and usually varies between 2 and 5. For our case, we assume that $\alpha = 2.3$ from experiments. In our experiment, β follows gamma(α_0, β_0), with $\alpha_0 = 2$ and $\beta_0 = 0.05$. The estimate of β varies in between 0.015 to 0.05 from frame to frame, because the pixel statistics change for different object positions, rotations, and illuminations.

The tracking of heavy occluded objects is usually difficult. Also, changes in rotation or illumination for the objects add more complexity to this problem. Consider that two objects [see Fig. 2(b)] move and rotate in between the occlusion and the background from frame to frame with varying illumination. In general, 2D algorithms may fail to track in this case. Experimental results with the 2D optimal object tracking algorithm presented in [17] are shown in Fig. 4. It can be seen that the performance using the 2D imaging approach is quite poor and the objects cannot be tracked. However, we show that our 3D tracking method performs reasonably well for this scene with changes in the orientation and illumination of the object. Reconstruction from elemental images for the first frame by using 3D computational method is shown in Fig. 5. For comparison, tracking with and without occlusion is performed. The performance for the first frame is shown in Fig. 6 (Media 1, Media 2, Media 3, and Media 4). Tracking examples of both cars are shown in Fig. 7 (Media 5 and Media 6). 3D tracking experiments are performed with varying orientation and illumination. Illumination is reduced by half in Fig. 7(b) for the tracking experiments, and car 2 and the opposite side of car 1 are tracked. Both cars are rotated in the tracking results in Fig. 7(c). Car 2 is rotated for tracking results in Fig. 7(d). Illumination is doubled, and car 1 is rotated by 135 degrees for tracking results in Fig. 7(e).

5. CONCLUSIONS

We have presented a Bayesian framework for tracking multiple objects in 3D space using a region tracking method based on statistical Bayesian formulation and 3D integral imaging. The proposed method is robust to partial occlusion and an unknown background scene, and it works with objects with unknown position, range, rotation, scale, and illumination. In the proposed tracking algorithm, the reconstructed pixel intensities of the background and the objects are assumed to follow Gaussian and gamma distributions, respectively. By assuming appropriate priors, posterior distributions of the background and the objects can be calculated. Multiobject tracking is achieved by maximizing the geodesic distance between the log-posteriors of the 3D reconstructed background and the objects. We have shown that statistical Bayesian formulation used with 3D integral imaging provides a promising technique for tracking objects in the 3D space.

ACKNOWLEDGMENTS

We wish to thank Dr. M. Daneshpanah and Prof. Dipak Dey for many useful discussions. This work was supported by the

Defense Advanced Research Projects Agency (DARPA) and by the Air Force Research Laboratory under FA8650-07-C-7740.

REFERENCES

1. G. Lippmann, "La photographic intégrale," C. R. Acad. Sci. **146**, 446–451 (1908).
2. A. Stern and B. Javidi, "3D image sensing, visualization, and processing using integral imaging," Proc. IEEE **94**, 591–608 (2006).
3. F. Okano, J. Arai, K. Mitani, and M. Okui, "Real-time integral imaging based on extremely high resolution video system," Proc. IEEE **94**, 490–501 (2006).
4. J. S. Jang and B. Javidi, "Three-dimensional synthetic aperture integral imaging," Opt. Lett. **27**, 1144–1146 (2002).
5. S. Hong and B. Javidi, "Improved resolution 3D object reconstruction using computational integral imaging with time multiplexing," Opt. Express **12**, 4579–4588 (2004).
6. B. Javidi, R. Ponce-Diaz, and S.-H. Hong, "Three-dimensional recognition of occluded objects by using computational integral imaging," Opt. Lett. **31**, 1106–1108 (2006).
7. M. Cho and B. Javidi, "Three-dimensional tracking of occluded objects using integral imaging," Opt. Lett. **33**, 2737–2739 (2008).
8. M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a handheld camera," Int. J. Comput. Vis. **59**, 207–232 (2004).
9. M. DaneshPanah and B. Javidi, "Segmentation of 3D holographic images using bivariate jointly distributed region snake," Opt. Express **14**, 5143–5153 (2006).
10. M. DaneshPanah and B. Javidi, "Tracking biological microorganisms in sequence of 3D holographic microscopy images," Opt. Express **15**, 10761–10766 (2007).
11. C. Chesnaud, V. Page, and P. Réfrégier, "Improvement in robustness of the statistically independent region snake-based segmentation method of target-shape tracking," Opt. Lett. **23**, 488–490 (1998).
12. A. Yilmaz, X. Li, and M. Shah, "Contour based object tracking with occlusion handling in video acquired using mobile cameras," IEEE Trans. Pattern Anal. Mach. Intell. **26**, 1531–1536 (2004).
13. M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: active contour models," in *Proceedings of the International Conference on Computer Vision* (IEEE, 1987), pp. 259–268.
14. T. Georgiou, "Distances and Riemannian metrics for spectral density functions," IEEE Trans. Signal Process. **55**, 3995–4003 (2007).
15. O. Germain and P. Réfrégier, "Optimal snake-based segmentation of a random luminance target on a spatially disjoint background," Opt. Lett. **21**, 1845–1847 (1996).
16. B. Javidi, P. Réfrégier, and P. Willett, "Optimum receiver design for pattern recognition with nonoverlapping signal and scene noise," Opt. Lett. **18**, 1660–1662 (1993).
17. F. Goudail and P. Réfrégier, "Optimal target tracking on image sequences with a deterministic background," J. Opt. Soc. Am. A **14**, 3197–3207 (1997).
18. C. Chesnaud, P. Réfrégier, and V. Boulet, "Statistical region snake-based segmentation adapted to different physical noise models," IEEE Trans. Pattern Anal. Mach. Intell. **21**, 1145–1157 (1999).
19. N. Mukhopadhyay, *Probability and Statistical Inference* (Marcel Dekker, 2000).
20. J. Sethian, *Level Set Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Material Sciences* (Cambridge University Press, 1999).