

Occluded Object Motion Estimation System based on Particle Filter with 3D Reconstruction

Kwang-Eun Ko, Junheong Park, Seung-Min Park, Jun-Yeup Kim, and Kwee-Bo Sim*

School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, 156-756, Republic of Korea

Abstract

This paper presents a method for occluded object based motion estimation and tracking system in dynamic image sequences using particle filter with 3D reconstruction. A unique characteristic of this study is its ability to cope with partial occlusion based continuous motion estimation using particle filter inspired from the mirror neuron system in human brain. To update a prior knowledge about the shape or motion of objects, firstly, fundamental 3D reconstruction based occlusion tracing method is applied and object landmarks are determined. And optical flow based motion vector is estimated from the movement of the landmarks. When arbitrary partial occlusions are occurred, the continuous motion of the hidden parts of object can be estimated by particle filter with optical flow. The resistance of the resulting estimation to partial occlusions enables the more accurate detection and handling of more severe occlusions.

Keywords : Occluded Object Tracking, Three-Dimensional Reconstruction, Particle Filter, Optical Flow

1. Introduction

Object tracking is an essential component of an intelligent vision system. Accurate and real-time object tracking will greatly improve the performance of object recognition, activity analysis and high-level behavior understanding [1]. However, the occlusion problem in multiple interacting object tracking has frequently occurred and methods to solve the problem have been previously presented. In essence, object tracking is the process of updating object attributes over time, which the attributes include position, motion, shape, and appearance. To suppress occlusion or noise and to achieve stable tracking, the attributes are smoothed by spatio-temporal filter [2].

Existing methods for smoothing the position and motion of object by Kalman filter or Monte-Carlo filters [5]. In both cases, however, the temporal filter has no effect on improving the object localization, the main task of the tracking. Wu [11] proposed a dynamic Bayesian network which accommodates an extra hidden process for partial occlusion handling. Although various approaches have been proposed in the literatures, the essence problem of occluded objects tracking in dynamic scene is still far from being completely solved.

To deal with a part of the problem, we are interested in the artificial imitation learning that is inspired by imitation learning

of sensory-motor in human brain called ‘mirror neuron system (MNS)’. MNS is discovered in area F5 of the monkey premotor cortex. It discharges both when the monkeys dose a particular action and when it observes another individual doing a similar action [3]. If mirror neurons are involved in action understanding, they should discharge also in conditions in which monkey does not see the occurring action but has sufficient clues to create a mental representation of what the experimenter does. The mirror neurons were tested in two basic conditions. In one, the monkey was shown a fully visible action directed toward an object (“full vision” condition). In the other, the monkey saw the same action but with its final, critical part hidden (“hidden” condition). Before each trial, the experimenter placed a piece of food behind the screen so that the monkey knew there was an object. Only those mirror neurons were studied that discharged to the observation of the final part of a grasping movement and/or to holding. More than half of the tested neurons discharged in the hidden condition. Out of them, about half did not show any difference in the response strength between the hidden and full vision conditions. The other half responded more strongly in the full-vision condition. One neuron showed a more pronounced response in the occluded condition than in full vision.

To model the function of MNS, firstly, it should be possible that reasoned about occluded part of the target object. Many researchers concentrated two-dimensional sight for the object tracking in occluded environment. However, it is hard to fully estimate the moving object in three-dimension. Based on our previous preliminary presentations [15], this paper presents a new updating method based on particle filter that satisfies the two qualities: accuracy and robustness.

The paper is structured as follows: Section 2 provides an overview of the related methods. Section 3 is the main section, which describes the particle filter based motion estimation for

Manuscript received Feb. 28, 2012; revised Mar. 14, 2012; accepted Mar. 15, 2012

* Corresponding author: kbsim@cau.ac.kr

This work was supported by the National Research Foundation of KOREA [NRF] grant funded by the KOREA government [MEST] [No. 2011-0029861]

© The Korean Institute of Intelligent Systems. All rights reserved.

It is the extended version of a paper awarded as a best presentation in ISIS2011.

occluded object. The experimental result of estimations is discussed in Section 4. Section 5 shows overall conclusion and future works.

2. Related Works

2.1. 3D Reconstruction

3D reconstruction enable tracing occluded part more easily, the detailed process is as follows. For this, original picture data have to be acquired by 3D-view stereo camera. And, the depth data calculated based on the voxel of the acquired picture. By using the depth information, we can proceed the preprocessing task. The preprocessing task is accomplished by finding correspondences between points of voxel that are seen by one imager and the same points as seen by the other imager [4]. Mathematically, radial and tangential lens distortion in original image is removed and draw undistorted images [5].

Adjust for the angles and distances between the cameras views (left, right), a process, called rectification, is performed. The outputs of this step are images that are row-aligned and rectified. Find the same features in the left and right camera views, a process known as correspondence. We applied SIFT (Scale-Invariant Feature Transform) algorithm to extract the same features. It is an algorithm widely used in computer vision to detect and describe local features in target images. The features are invariant to image scale and rotation, and are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, additional noises, and change in illumination [6]. The features are highly distinctive in the sense that a single feature can be correctly matched with high probability against a large database of features from many images [7]. It detects key points using a cascade filtering approach that uses efficient algorithms to identify candidate locations that are then examined in further detail. The first stage of key point detection is to identify locations and scales that can be repeatable assigned under differing views of the same object. Detecting locations that are invariant to scale change of the image can be accomplished by searching for stable features across all possible scales, using a continuous function of scale known as scale space. The output of this step is a disparity map, where the disparities are the differences in x -coordinates on the image planes of the same feature viewed in the left and right cameras. If we know the geometric arrangement of the cameras, then we can turn the disparity map into distances by triangulation. This step is called re-projection, and the output is a depth map.

We assumed two equivalent camera modules that configures stereo camera. If the each camera module is not equivalent, color value of image is different according to the character of each camera as well as it has difference 'Focal Length'. Then, we should do revision. For this, we must be used each algorithm that is suitable for each camera system. Therefore, we used the same camera. Next, we settled the camera has same 'Epipolar Line' when we installed these. If the 'Epipolar

Line' not coincides, the speed of system has slowed because the system used overfull equation for revision error between cameras.

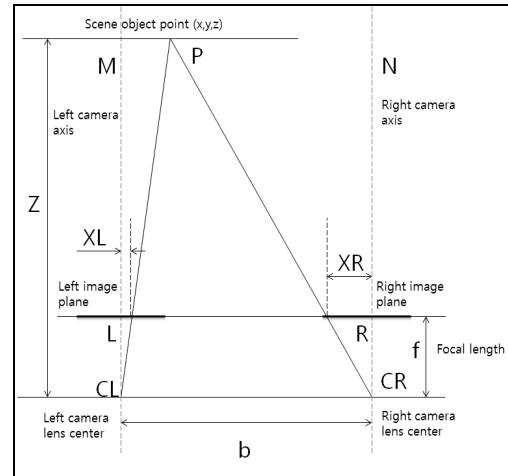


Fig. 1. Principle of 3D data extraction.

Now, we can see the method for extracting actual three dimensions data set. The point P exists in field at the Fig. 1 and the point P is expressed one point in the image on right camera and left camera. XL and XR mean x -coordinate of P in the image on two cameras and b is distance between two cameras. F means 'Focal Length'. Using this, we solved proportional expression according to resemblance of triangle and then we got the following (1) that obtained value of z that the distance between two camera modules.

$$z = \frac{\text{Camera Distance}(b) \times \text{Focal Length}(f)}{|XL - XR|} \quad (1)$$

If find value of z , y and x are calculated easy by trigonometry with Focal Length, value of z , and x and y values of image.

2.2. Optical Flows

The last stage for reconstructing 3D images is building the object history. The building of history is maintaining 3D data rebuilt in the prior stage. If the object turned in sight of stereo camera, it maintain 3D data through traced movement of object [8]. For this, we applied optical flow that is widely used in many research associated with the tracing of object [9]. Optical flow based motion vector is an important information about the movement of object. Once the optical flow field is computed accurately, the measurement of image velocity can be used widely in many tasks in computer vision area. In this paper, we applied Lucas-Kanade (LK) optical flow algorithm. LK algorithm is also measure motion field, i.e. information of motion, using difference between brightness values of picture on the assumption that equal the speed in the local field had characteristic of picture. LK algorithm is method that calculated motion field about local field. It has advantage that

is accurate and had a strong resistance to noise. It has basic character as in the following.

In the picture, pixel is expressed $\mathbf{u} = (u, v)$ and time t is expressed $I(\mathbf{u}, t)$. Pixel \mathbf{u} is moved to $\mathbf{u}' = F(\mathbf{u}, \mu)$ by motion vector when the time $t+1$. In this time, the \mathbf{u} is motion vector and the $F(\mathbf{u}, \mu)$ is motion model. If surrounding light is fixed, (2) comes into existence because pixel coordinate is moving in the picture but value of brightness is not changed

$$I(F(\mathbf{u}, \mu), t+1) = I(\mathbf{u}, t) \quad (2)$$

At this time, LK algorithm calculated value of motion vector \mathbf{u} minimizing value of brightness variation square like (3).

$$\min E(\mu) = \sum_{\mathbf{u} \in \Omega} (I(F(\mathbf{u}, \mu), t+1) - I(\mathbf{u}, t))^2 \quad (3)$$

In the (3), Ω is measured object region that calculated using pixel existed in the Ω for getting value of motion vector.

We mention 3D reconstruction through stereo camera. And we refer estimation and tracing of moving object using optical flow and particle filter. Fig. 2 shows the flowchart that is proposed in this paper.

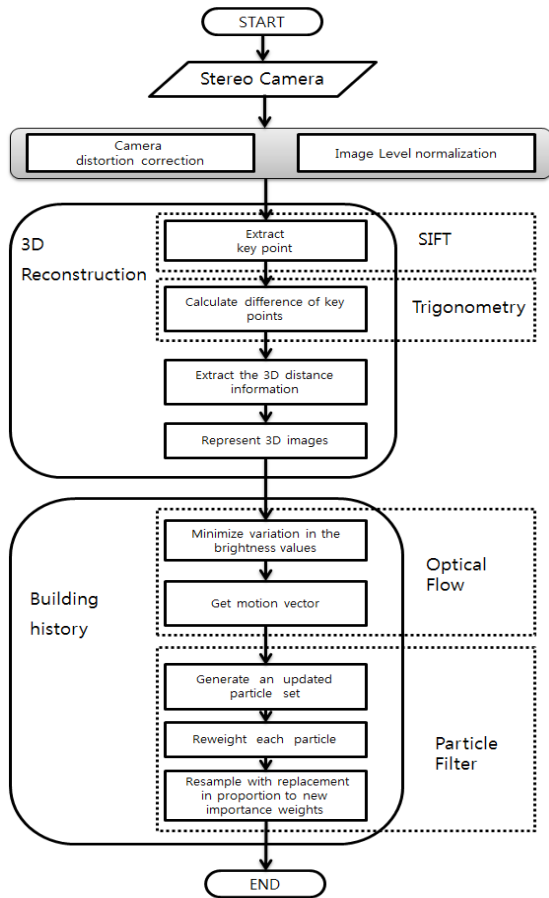


Fig. 2. Flowchart of 3D reconstruction and building history.

The present direction was grasped using understanding of moving object by picture data reconstructed as 3D through optical flow. And the history about former 3D data built and we traced motion using particle filter persistently. If the part was visible in the past but invisible in the present we getting the data of the part based on the history that was built former stage and using this data in 3D reconstruction entire process simply.

3. Proposed Method

3.1. Particle Filter

First introduced in 1960, the Kalman filter has risen to great prominence in a wide variety of signal processing contexts [5]. The Kalman filter models a single hypothesis. It will be possible—given a history of measurements of a system—to build a model for the state of the system that maximizes the a posteriori probability of those previous measurements. Because the underlying model of the probability distribution for that hypothesis is uni-modal Gaussian, it is not possible to represent multiple hypotheses simultaneously using the Kalman filter [10]. A somewhat more advanced technique known as the condensation algorithm, which is based on a broader class of estimators called particle filters, will allow us to address this issue [11].

The operation principle of the particle filter is very simple [12]. Particle Filter is the most widely used algorithm on tracing object [13]. Particle Filter is one of the Prediction Technique that Sequential Monte Carlo method based on simulation. And it is suitable for non-Linear, non-Gaussian system. The object of particle filter is predicted income information with only observed value had error. The motion principle of particle filter is so simple. We putted several pressures on appropriate proposed probability distribution in the system. Next, particle filter is synthesized all data and was supposed information of system. Like method of Bayesian filtering, particle filter observed posterior probability of Markov process that used (4) repeat.

$$p(X_t | Y_{1:t}) = \alpha p(Y_t | X_t) \int_{X_{t-1}} p(X_t | X_{t-1}) p(X_{t-1} | Y_{1:t-1}) \quad (4)$$

Expect basically cases, integral of (4) hadn't closed from solution. Therefore, (4) is used the set $\{(X_t^{(i)}, Y_t^{(i)})\}_{i=1, \dots, N}$ of weighted samples (particles) thereby making approximate value. In this equation, $X_t^{(i)}$ means particle and $w_t^{(i)}$ means weight values applied each particle. Current object state $P(X_t^{(i)} | Y_{1:t})$ is expressed by using particle $X_t^{(i)}$ and weight values $w_t^{(i)}$. It is (5). In the (5), δ is means Dirac-delta function

$$p(X_t | Y_{1:t}) \cong \sum_{i=1}^N w_t^{(i)} \delta(X_t - X_t^{(i)}) \quad (5)$$

After (4) changed Equation concerning previous object state $P(X_{t-1}^{(i)} | Y_{1:t-1})$, the changed equation is substituted for (4) and arranged the equation. We could get (6).

$$p(X_t | Y_{1:t}) \cong \alpha p(Y_t | X_t) \sum_{i=1}^N w_{t-1}^{(i)} p(X_t | X_{t-1}^{(i)}) \quad (6)$$

In order to maintain the filtering distribution of (6), the fact that N particles set are expressed filtering distribution in the prior stage is supposed. It is $P(X_{t-1}^{(i)} | Y_{1:t-1}) \cong \{(X_{t-1}^{(i)}, w_{t-1}^{(i)})\}_{i=1, \dots, N}$, new N particles set $X_t^{(i)} |_{i=1, \dots, N}$ are sampled by the proposal distribution $X_t^{(i)} \cong q(X_t | X_{0:t-1}, Y_{1:t}) |_{i=1, \dots, N}$ and importance weights about this particle set calculated according to (7).

$$w_t^{(i)} \propto w_{t-1}^{(i)} \frac{p(Y_t | X_t^{(i)}) p(X_t^{(i)} | X_{t-1}^{(i)})}{q(X_t^{(i)} | X_{0:t-1}, Y_{1:t})} \quad (7)$$

The last stage of particle filter is that particle goes through resampling process proportionally to importance weight.

3.2. Particle Filter based Object Motion Estimation

The data flowchart diagram for estimation iteration is shown in Fig. 2. We have implemented the proposed algorithm for two kinds of features: *a*) motion vector \mathbf{u} as in (3) and *b*) image intensity f^I that is introduced in follows:

$$f^I = \frac{1}{3}[R + G + B] \quad (8)$$

A motion vector \mathbf{u} represents the temporal characteristics of camera motion. The particle filter based motion tracking algorithm is described as follows:

- 1) From the previously estimated mean state vector, run the LK optical flow. Compute the motion vectors, and perform the resample before drawing the new sample set.
- 2) Resample from the old sample set and choosing a sample according to its post probability (likelihood).
- 3) By using the sampling scheme as follows:

$$X_t^{(i)} = X_{t-1}^{(i)} + \Delta X_t + \varepsilon_t^{(i)}, \quad \varepsilon_t^{(i)} \sim N(0, \sigma^2) \quad (9)$$

- 4) Weight the samples according to following equation (7).
- 5) The estimated state vector at time t is then given by the weighted average of all the particles at time t .

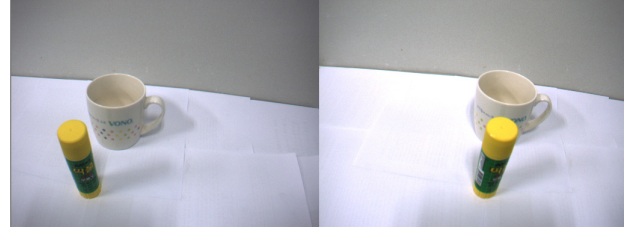
4. Simulation

4.1. Experimental Condition

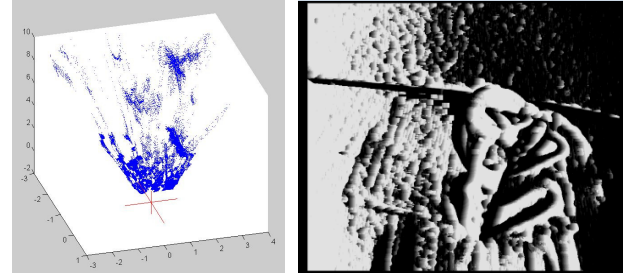
To simulate the proposed method, we applied multiple objects for tracking and estimation in 25 image sequence with 640x480 at 480 frames per second acquired from Bumblebee 2 stereo camera. The partial occlusion of object is included in each image sequence. The simulation is implemented on standard PC environment (Intel Core i5-2500 3.30GHz). We deliberately selected image sequences taken under different conditions, such as illumination, background, and so on. The complete occlusion is declared when the target object occupies less than 25 percent of the region found by the algorithm.

4.2. 3D reconstruction using Stereo Camera

Fig. 3 illustrates an example of the 3D reconstruction of partially occluded object (mug). By using the reconstruction process, we can acquire the depth information of target object in the original image and reconstruct the 3D image data.



(a) Target Object Images from Stereo Camera



(b) Simulated Depth Data (left), Reconstruction Image (right)

Fig. 3. Simulation result of 3D reconstruction.

The tracking results for the proposed method and comparison algorithms were collected in Table 1.

Table 1. Statistics of the Estimation Results for Test Data Set.

Algorithm	3D reconstruction + Particle Filter	Kalman Filter	Template Matching
Loss Rate	9/25	12/25	10/25
False Occlusions	6	4	10

5. Conclusion

In this paper, 3D reconstruction method is supposed for estimation of motion about invisible part among the character of mirror neuron system. This method is that obtained depth information using stereo camera and reconstructing the three dimensions image. And then, history of the features is built using optical flow and the motion vector and image intensity are calculated as motion feature. Particle filter-based motion estimation algorithm is applied to trace the occluded target object.

By using the method, if we do intention recognition by motion, improvement of recognition improve accuracy is expected. In future work, we use hidden Markov model that is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved states as classification algorithm.

References

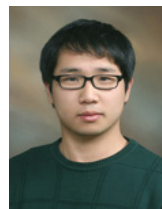
- [1] T. Yang, Q. Pan, J. Li, S.Z. Li, Real-time multiple objects tracking with occlusion handling in dynamic scenes, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005 (CVPR 2005), vol. 1, pp. 970-975, June 20-25, 2005.
- [2] H. T. Nguyen, A. W. M. Smeulders, "Fast occluded object tracking by a robust appearance filter," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp. 1099-1104, August 2004.
- [3] G. Rizzolatti and L. Craighero, "The mirror-neuron system.," Annual review of neuroscience, vol. 27, pp. 169-92, Jan. 2004.
- [4] D.-hun Lee, D.-youn Um, and H. Kang, "A Study of the 3D-Reconstruction of indoor using Stereo Camera System," Journal of Korean Institute of Intelligent Systems, vol. 15, no. 1, pp. 42-47, 2005.
- [5] A. Blake, R. Curwen, and A. Zisserman, "A framework for spatio-temporal control in the tracking of visual contour," International Journal of Computer Vision, vol.11, no. 2, pp. 127-145, 1993.
- [6] D. G. Lowe, "Object recognition from local scale-invariant features," Proc. the Seventh IEEE International Conference on Computer Vision, pp. 1150-1157 vol.2, 1999.
- [7] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, Nov. 2004.
- [8] S. Beauchemin, J. Barron, and D. Fleet, "Systems and experiment performance of optical flow techniques," Int. J. Comput. Vision, vol. 12, no. 1, pp. 43-77, 1994.
- [9] J. Chun and G. Shin, "Realtime Facial Expression Recognition from Video Sequences Using Optical Flow and Expression HMM," Journal of Korean Society for Internet Information, vol. 12, no. 3, pp. 55-70, 2011.
- [10] P. Hargrave, "A tutorial introduction to Kalman filtering," Kalman Filters: Introduction, Applications and, pp. 1-6, 1989.
- [11] Y. Wu, T. Yu, G. Hua, "Tracking appearance with occlusions" Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 789-795, 2003.
- [12] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," IEEE Transactions on Signal Processing, vol. 50, no. 2, pp. 174-188, 2002.
- [13] H. B. Kim, K. E. Ko, J. S. Kang, K.-B. Sim, "Specified Object Tracking in an Environment of Multiple Moving Objects using Particle Filter," Journal of Korean Institute of Intelligent Systems, vol. 21, no. 1, pp. 106-111, 2011.
- [14] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, vol. 77, no. 2, pp. 257-286, 1989.
- [15] J. H. Park, S. M. Park, K. E. Ko, K. B. Sim, "Occluded object motion tracking method based on combination of 3D reconstruction and optical flow," International Symposium on Advanced Intelligent Systems, Suwon, Korea, Oct. 2011.
- [16] N. Bouaynaya, D. Schonfeld, "A complete system for head tracking using motion-based particle filter and randomly perturbed active contour," Proc. SPIE, vol. 5685, pp.864 - 873 , 2005.



Kwang-Eun Ko

He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2007. He is currently Ph. D course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Machine Learning, Brain-Computer Interface, Gesture Recognition, etc.

E-mail : kkeun@cau.ac.kr



Junheong Park

He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2011. He is currently Master course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Machine Learning, Brain-Computer Interface, Gesture Recognition, etc.

E-mail : antiline@cau.ac.kr



Seung-Min Park

He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2010. He is currently Ph. D course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Pattern Recognition, Brain Computer Interface, Evolutionary Computation, etc.

E-mail : sminpark@cau.ac.kr



Jun-Yeup Kim

He received the B.S. degree from the Department of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Korea, in 2012. He is currently Master course in the School of Electrical and Electronics Engineering at Chung-Ang University. His research interests include Machine Learning, Brain-Computer Interface, Gesture Recognition, etc.

E-mail : jy915@cau.ac.kr



Kwee-Bo Sim

He received his B.S. and M.S. degrees in the Department of Electronics Engineering from Chung-Ang University, Korea, in 1984 and 1986 respectively, and Ph.D. degree in the Department of Electronics Engineering from The University of Tokyo, Japan, in 1990. Since 1991, he is currently a Professor. His research interests include artificial life, emotion recognition, ubiquitous intelligent robot, intelligent system, computational intelligence, intelligent home and home network, ubiquitous computing and Sense Network, adaptation and machine learning algorithms, neural network, fuzzy system, evolutionary computation, multi-agent and distributed autonomous robotic system, artificial immune system, evolvable hardware and embedded system etc. He is a member of IEEE, SICE, RSJ, KITE, KIEE, KIIS, and ICROS Fellow.

Tel : +82-2-820-5319

Fax : +82-2-817-0553

E-mail : kbsim@cau.ac.kr