H–008

Occlusion Handling of Multiple Objects in Indoor Depth-based Surveillance System

Houari Sabirin† Sei Naito†

1． Introduction

With the availability of less costly depth sensors recently object detection and tracking using RGB-D or 3D range imaging cameras has been commonly practiced. Such cameras have been cited for their advantages in using the availability depth data to enable detection of an image that cannot be perceived by a visible light sensor as in an RGB camera. This feature would enable reliable object recognition when color information is insufficient. Depth information can provide additional measurement to distinguish objects by the distance value of each pixel of the object, by which color information may lead to incorrect perception. For example, in a dark room a depth camera would still be able to detect moving objects compared with a conventional camera. On the other hand, because a depth camera does not perceive color signals, it has the advantage of protecting the privacy of the monitored people in a public area by not exposing their faces.

In this paper we propose object detection and tracking method by optimizing the utilization of only depth information without RGB signals to identify moving objects. Here the identity of the objects can be preserved and correctly assigned during occlusion even when the objects overlap with each other as in conventional (i.e., RGB camera) object detection and tracking. Depth-only image analysis using such cameras has been presented in [1], [2]-[4] to perform single object tracking to track selected single static objects (such as a face, cup, toys, etc.) and correctly identify the object even when the object is occluded with moving subjects (such as moving hands or books) [1], to track single objects without occlusions [3], [4] and to track multiple objects with occlusion detection but no identification during the occlusion [2].

2． Proposed Method

Our method utilizes two network-based static range sensor cameras with environment settings as illustrated in Fig. 1. Both cameras are located at each end of the room, roughly 8.5 meters apart, and 2.6 meters from the floor with a tilt angle of around 45 degrees. The coverage distance of each camera is set to 6 meters with around 70 degrees of field of view. These settings will produce an overlapping area from both cameras at the center of the room, which is useful for maintaining the object's displacement from one camera to another.

Our method begins by creating a background model based on [5] to be utilized in the background subtraction process to produce object masks from each camera. These masks are then projected onto a 2D plane, and combined into a single *processing frame* where object detection and tracking procedures will be performed. The background subtraction process is performed by binarizing the depth data in a frame as

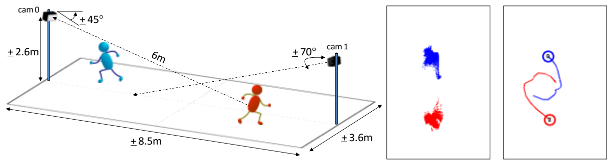

Fig. 1 Camera positions in room for the experiments (left) and processing plane of detected object and its tracking result (right).

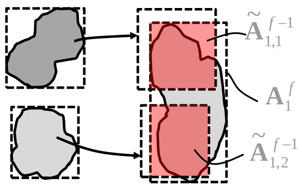

Fig. 2 An example of occlusion determination of two objects between the previous frame (left) and the current frame (right)

$$D(i,j)=\begin{cases}0 & ,\min(\hat{\mathbf{d}})-\alpha\max(\Delta\hat{\mathbf{d}})< \\ & \quad D(i,j)<\max(\hat{\mathbf{d}})-\beta\max(\Delta\hat{\mathbf{d}}) \\ 255 & ,\textit{otherwise}\end{cases} \qquad (1)$$

where $D(i,j)$ is a depth value at position i,j in a frame. Depth set $\hat{\mathbf{d}}$ is the depth value at position i,j in each of reference frames that satisfy $\left|D_f(i,j)-\lambda(i,j)\right|\le 2\sigma(i,j)$ where λ and σ are the median and standard deviations of the depth value at position i,j of the reference frames, and the depth value difference between two consecutive frames $\Delta\tilde{\mathbf{d}}=\left|D_f(i,j)-D_{f-1}(i,j)\right|$. Constant values of α and β are set to 1.0 and 3.0 respectively based on experiments.

Next, to automatically detect occlusions in the processing frame, overlap checking is performed by calculating the size of the overlapping area of the objects masks in the current and previous frame. If an occlusion is detected, object separation is then performed to split the occluded objects into the corresponding objects in the previous frame, before the occlusion has occurred. Let $\mathbf{A}_1^{f-1}$ as the area of *object* 1, $\mathbf{A}_2^{f-1}$ as the area of *object* 2 in previous frame, respectively; and $\mathbf{A}_1^f$ as the area of an object in current frame. Suppose the positions of *object* 1 and *object* 2 in previous frame correspond to the position of an object in current frame such that their areas $\mathbf{A}_1^{f-1}$ and $\mathbf{A}_2^{f-1}$ are intersect with $\mathbf{A}_1^f$ as illustrated in Fig. 2. To determine that the overlap indicates occlusion, two conditions apply: the intersect areas of $\mathbf{A}_1^{f-1}$ and $\mathbf{A}_2^{f-1}$ with $\mathbf{A}_1^f$, denoted by $\tilde{\mathbf{A}}_{1,1}^{f-1}$ and $\tilde{\mathbf{A}}_{1,2}^{f-1}$ respectively, shall be larger than a given threshold; and at least two areas are overlapping between the current object and any objects in the previous frame.

In case of occlusions, the separation step plays an important role in simplifying object tracking where initial areas of each occluded object need to be distinguished in order to enable in-occlusion tracking. A separation method based on the spatial

† KDDI R&D Laboratories, Inc.

information of the occluded objects in the previous frame and the

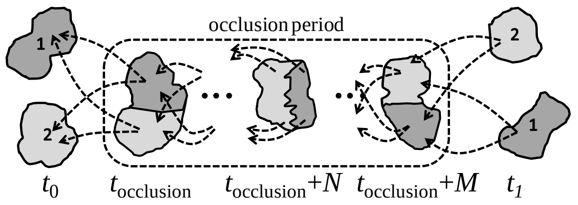

Fig. 3 Illustration of object tracking scheme during occlusion.

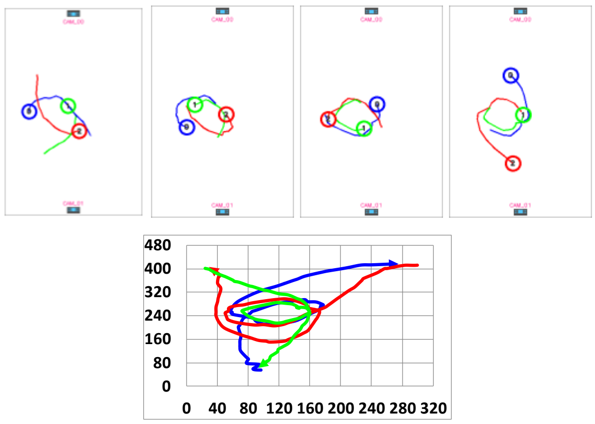

Fig. 4 An example tracking results of three objects: (above) 2D projection during occlusion captured every 5 frames, (bottom) plot of their trajectories (vertical axis is the frame height; horizontal axis is the frame width).

depth information of the merged area in the current frame is performed.

The separation method virtually removes all occlusion occurrences, and tracking can be simply performed by similarity matching using eq. (2). As illustrated in Fig. 3, suppose an occlusion has occurred between frame t_0 and t_1. At the beginning of the occlusion period in $t_{occlusion}$, the occluded objects are already separated such that by using similarity matching, individual attributes of each occluded object can be determined. Since overlap checking and separation will be continually conducted during the occlusion period, each part of the occluded object that has been separated will be assigned similarity matching with the previously separated objects. Thus their individual identities can be retained. As a result, when the occlusion has finished, the identity of each object in t_1 can also be correctly determined.

After separation has been completed, object tracking during occlusion can be performed. In brief, similarity matching similar to [2] is utilized in our method with respect to spatial information of the objects. Object tracking can be handled by performing similarity matching between the attributes of two objects in two consecutive frames. More specifically, let $\mathbf{P}_k^f = \{i, j\}_k^f$ be the position of the k-th object at i,j; thus the k-th object in frame f is assigned the same identity as the m-th object in frame f-1 if the following rule is satisfied

$$k_{match} = \arg\min_{m}\left(\left\|\mathbf{P}_k^f - \mathbf{P}_m^{f-1}\right\|\right). \qquad (2)$$

3. Experimental Results

We construct our test sequences in VGA resolution (640×480) with various occlusion events prepared from the experimental environment in Fig. 1 to evaluate the proposed method. The experimental objects are people walking at normal speed. The proposed method emphasizes the separation of occluded objects for occlusion handling, with the object detection and tracking data being combined in one single processing frame. Some examples of identification and trajectory labeling resulting from our proposed method are shown in Fig. 4. Here, three objects are walking around each other. The images are captured every five frames to show the accuracy of the resulting trajectories during occlusions. The tracking results show that all objects can be consistently tracked during occlusion, and consequently their identification can still be correctly retained after occlusion.

Some failure cases occurred when the depth masks are fully overlapped with each other. In such case, the mask of the object covered by the other's mask is lost, and the tracking result cannot be displayed. In short and narrow occlusion, when the mask is uncovered, its identity can be retained by checking the last missing objects during the tracking. However, when the object is missing for a longer time period and their last position is too different from its last position, incorrect identification may occur[1].

4. Conclusions

We have shown in this paper a novel method to identify and track moving objects during occlusion in a depth-based object detection and tracking system, where no color signals are available. By firstly separating the occluded objects based on their depth information, the tracking task can be made simpler where similarity matching can be done for each individually separated object.

References

[1] S.-C. Shen, W.-L. Zheng, and B.-L. Lu, "Online Object Tracking Based on Depth Image with Sparse Coding," Neural Information Processing, pp. 234-231, 2014.

[2] H. Sabirin, H. Sankoh, and S. Naito, "Utilizing attributed graph representation in object detection and tracking for indoor range sensor surveillance cameras," IEICE Transactions on information and Systems, vol. 98, no. 12, pp. 2299-2307, 2015.

[3] U. Reiser, and J. Kubacki, "Using a 3D Time-Of-Flight Range Camera for Visual Tracking," Symposium on Intelligent Autonomous Vehicles, 2007.

[4] L. Jia, and R.J. Radke, "Using Time-of-Flight Measurements for Privacy-Preserving Tracking in a Smart Room," IEEE Trans. on Industrial Informatics, vol. 10, no. 1, pp. 689-696, 2013.

[5] S.H. Cho, K. Bae, K.M. Kyung, S. Jeong, and T.C. Kim, "Background Subtraction Based Object Extraction for Time-of-Flight Sensor," in *Proc. IEEE 2nd Global Conf. on Consumer Electronics*, pp. 48-89, 2013.

[1] The results of our tracking can be viewed online at https://youtu.be/HqnMci-hfTA