A Layered-based Fusion-based Approach to Detect and Track the Movements of Pedestrians through Partially Occluded Situations

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
A Layered-based Fusion-based Approach to Detect and Track the Movements of Pedestrians through Partially Occluded Situations

Yajun Fang, Sumio Yokomitsu, Berthold K.P. Horn, Ichiro Masaki

Abstract—To obtain perception abilities, conventional methods independently detect static and dynamic obstacles, and estimate their related information, which is not quite reliable and computationally heavy. We propose a fusion-based and layered-based approach to systematically detect dynamic obstacles and obtain their location and timing information. The layered-based concept helps us to first search pedestrians in horizontal dimension based on transitional peaks in the defined projection-curves, and then search in vertical dimension. Converting a typical 2D search problem into two 1D search problems significantly decreases the computational load. The fusion-based obstacle detection fuses the information from initial segmentation and dynamic tracking model to avoid complicated tracking schemes. The methodologies take advantage of connection between different information, and increase the accuracy and reliability of obstacle segmentation and tracking. The search mechanism works for both visible and infrared sequences, and is specifically effective to track the movements of pedestrians in complicated environments such as human intersecting and conclusion, thus improving environment understanding abilities and driving safety.

I. INTRODUCTION

In order to enhance the safety of intelligent transportation systems, a large amount of surveillance cameras are mounted for the purpose of intelligent monitoring and automatic environment and event understanding. Recently intelligent intersection surveillance has attracted lots of attention and it is expected to understand the human behavior of passing pedestrians which involves automatic human detection, identification and activity understanding. Such information can help to provide early warning information for vehicles and for intelligent management of transportation systems.

For automatic environment understanding, it is required to identify interested objects and to understand their static and dynamic information. Typically two steps are involved: initial segmentation in the first few frames, which is typically very complicated, and human tracking in the subsequent frames to detect the identified foregrounds in the first step, which is more computationally efficient than the first step.

Traditional initial segmentation usually depends on special assumptions and is challenged by unpredictable background, specifically with non-rigid objects, such as trees, or constantly-changing noises. Researchers often take advantage of visual features, such as obstacles’ motion, the symmetry of obstacles’ regions and asymmetry of background, etc., to differentiate foregrounds from backgrounds. Segmentation and tracking algorithms are also challenged by the significant variation of image positions and sizes of objects between successive video frames. Thus the accuracy and reliability of current human detection and identification algorithms are limited in many difficult scenarios, for example, when there is occlusion, when people wear clothes close to background’s color, when there are unexpected dynamic movements, etc. Conventional methods independently detect information for obstacles, and estimate their dynamic current historic information. Each process is computationally heavy and is limited in its reliability. We need to develop a new algorithm with higher performance and fewer assumptions than traditional ones.

II. FUSION-BASED LAYERED-BASED DYNAMIC TRACKING OF PEDESTRIANS

We proposed fusion-based and layered-based methodologies to systematically detect dynamic obstacles and obtain their dynamic information for visible and infrared sequences.

Sensor fusion is a common technology to improve the detection performance by combining information from different resources. The technology can be divided into three categories: decision-level fusion, data-level fusion, and feature-level fusion [1] [2]. Decision-level fusion takes all sensors’ decisions into account when making final decisions at a higher level. Data-level fusion combines raw data from different resources statistically, for example, through voting technique, Bayesian, and Dempster Shafer, etc., which require all data in similar format. In contrast, feature-level fusion takes advantage of specific information from one sensor and uses the information in another sensor’s process, which needs special algorithms to incorporate any additional information.

We proposed feature-level fusion-based obstacle detection methods. The proposed methodologies take advantage of connection between different information to improve obstacle detection and perception abilities. The extra information can help to convert a complex segmentation task into several simpler ones so that the time-consuming full-image searching and segmentation can be avoided. Layered-based separation is one of key components for whole systems, which is to separate an original image into several layers on which segmentation is simpler than in the original image.

Here below we first introduce how the fusion-based and layered-based principle is applied to detect and track the locations of obstacles in visible images from stereo camera [5][6][7] for daytime and in infrared images from single camera for nighttime [8][9]. The we will propose how to detect
and track the locations of pedestrians in visible images from a single camera.

A. Layered-based Image Separation for Stereo Visible Images and Infrared Images

In our previous work, we have proposed two methods to separate one image into several layers in which segmenting interested targets is much easier than in original images.

The first category of layers that we defined are distance-based image-layers [5] [6][7]. As shown in figure (1), an edge map is divided into several distance-based layers with the same size as the original images. In these layers, obstacles at specific distance-ranges stand out and can be identified easily. The method can also be used to remove the noises from non-rigid background objects that are far away as shown in the second example in figure (1).

The second category of layers are several vertical stripes within infrared images in which we can detect potential pedestrians for night vision [8][9] as shown in figure (2). The proposed strategy simplifies the original two-dimensional full-image segmentation into two one-dimensional segmentation by splitting an image with a size of $n_{row} \times n_{col}$ into several vertical stripes with a size of $n_{row} \times n_i$, $n_i < n_{col}$. Within these vertical stripes, images of human beings can be further segmented based on either brightness or bodylines. This is called “Horizontal segmentation first, vertical segmentation second” scheme.

For night vision, the key to the above automatic human-size estimation [8][9] is to split infrared image horizontally based on bright-pixel-vertical-projection curves, i.e., the number of bright-pixels in image columns versus their corresponding horizontal positions [9]. As shown in figure (2)(a) and (b), the horizontal locations and width of projection waves correspond to pedestrians and are robust to parameter choices.

B. Initial Horizontal Segmentation for Visible Images from Single Camera

In this section, we will introduce how to apply the idea of “Horizontal segmentation first, vertical segmentation second” to pedestrian detection and tracking in visible images from a single camera. “Vertical-image-striped separation” is based on the vertical projection of difference-images among consecutive frames. The method can be used for difference-image-based moving object detection.

Figure (3) explains the principle of horizontal segmentation based on the vertical projection of differential images. Figure (3)(a)(b) shows a simple example when ignoring the impact of backgrounds, i.e., the vertical projection of an image containing only foregrounds. Since the typical vertical projection of human being within differential images is in the shape of sharp triangle spikes, we use a triangle as an example of vertical projection of foregrounds. The three triangles $V_{-1}$, $V$, and $V_{+1}$ in figure (3)(a) respectively correspond to the vertical projection of same foreground at three continuous frames. The top two figures in figure (3)(b) show the vertical
projection of the differential image $V - V_{-1}$ and $V - V_{+1}$. The intersection of these two figures will provide the accurate locations and widths of the corresponding foregrounds. The principle applies to situations with general backgrounds. The projection of differential images will remain the same for the general situations with backgrounds since the impact of backgrounds on the vertical projection can be removed when computing differential images and their vertical projection-curves. The top image in the figure (3)(c) is the intersection of two differential images based on three continuous frames in figure (3)(e). The vertical intersection in figure (3)(c) is independent of backgrounds and provides the accurate location of two pedestrians. Specifically the peaks of the vertical projection correspond to the horizontal locations of their heads.

The above conclusion also holds when there are occlusions. As shown in figure (4), the vertical projections of the differential images provide us horizontal locations of interested pedestrians. Even if the horizontal locations of pedestrians do not necessarily correspond to global peaks, they do correspond to local transition-peaks, which helps to narrow the searching range of interested objects.

In sum, our “projection-based, horizontal first, vertical next” segmentation algorithm separates an image into several vertical stripes based on the following steps:

1) Compute the intersection of two continuous differential images from three consecutive images as our defined differential-images.

2) Calculate vertical projection-curves of bright pixels for the above defined differential images.

3) Detect triangle spikes to capture the big transitions in the obtained vertical projection-curves and merge neighbor spike when necessary. We adaptively adjust minimum wave-height threshold based on wave-height in previous frames in order not to miss objects.

Thus, we have obtained the initial horizontal segmentation. For non-occlusion cases, the horizontal locations of the triangle spikes provide the ideal width and horizontal locations of potential moving objects. The horizontal locations of the triangle peaks provide the horizontal head locations of potential moving pedestrians as shown in figure (3).

For occlusion cases, the horizontal locations of the triangle spikes still provide the horizontal boundaries of two potential moving objects. And the horizontal locations of local transition peaks on projection-curves provide the horizontal head locations of potential moving pedestrians as shown in figure (4).

Figure (5) shows the changing trends of vertical projection-curves of differential images before and after two people intersect. The correspondence between these vertical projection-curves and segmentation images shows relationship between a series of pedestrian locations and their vertical projection computed from defined differential images. Two individual peaks in the projection-curves gradually move closer, merge when occlusion happens, and finally depart again. The peaks of vertical projection correspond to the horizontal locations of pedestrians during the whole process. The series of vertical projection-curves provide a good foundation for initial segmentation.

![Fig. 3. Principle of horizontal segmentation based on the intersection of two continuous differential images (a) The projection of three continuous sample images with pedestrians (b) To recover the object width based on the intersection of two continuous differential images. (c) Horizontal segmentation based on the projection of the special defined differential images (e1)-(e3) Three continuous frames that produces the top image in (c)](image)

![Fig. 4. An example that the horizontal location of pedestrians correspond to transition peaks of the projection. Dotted lines correspond to the boundaries of intersection regions. Solid lines correspond to the locations of pedestrians’ heads. We notice that the horizontal locations of people’s head are exactly at the transitional peaks of the projection-curves, and the horizontal locations of merged regions at accurately determined by the peak boundaries of projection-curves.](image)

**C. Initial Vertical Segmentation within Horizontal Segmented Stripes**

For initial segmentation, the vertical positions of human beings can be directly obtained based on vertical distribution of edge pixels within differential horizontally segmented stripes. After detecting initial locations for pedestrians, we obtain the human body trunk templates for the purpose of future template tracking of pedestrians. Though people change their poses during walking, the corresponding body trunks change little between two continuous frame and can be used to match pedestrians among different frames.
For update step, the measured horizontal locations and speeds of previously identified pedestrians are used to update the gain, measurement, and error covariance of KF models as shown in equation (3).

The states of each KF model include the horizontal position \( \text{pos}_x \) and horizontal speed \( v_x \) for each pedestrian, i.e., \( \mathbf{x} = [\text{pos}_x \ v_x] \). We set \( B = [0; 0] \), \( H = [1 \ 0] \), \( A = [1 \ 1; 0 \ 1] \) since \( \text{pos}_x[k] = \text{pos}_x[k-1] + v_x[k-1] \) and \( v_x[k] = v_x[k-1] \).

Kalman filtering involves the step of estimation and update. For estimation step, we estimate horizontal locations and speeds for humans, and the error covariance in the next frame based on historical data as shown in equation (2).

\[
\begin{align*}
\hat{x}_k &= A\hat{x}_{k-1} + Bu_{k-1} \\
\hat{P}_k &= AP_{k-1}A^T + Q
\end{align*}
\] (2)

For update step, the measured horizontal locations and speeds of previously identified pedestrians are used to update the gain, measurement, and error covariance of KF models as shown in equation (3).

\[
\begin{align*}
K_k &= P_k^{-1}H^T(HP_k^{-1}H^T + R)^{-1} \\
\hat{x}_k &= \hat{x}_k + K_k(z_k - H\hat{x}_k) \\
P_k &= (I - K_kH)P_k^{-1}
\end{align*}
\] (3)

Different from traditional Kalman filtering where predicted results from equation (2) are only intermediate results, we actively take advantages of the predicted horizontal location and speed of pedestrians. The information is sent to a fusion-based tracking block to improve its reliability. As shown in the flowchart in figure (6)(a), fusion-based tracking block receives information from initial segmentation and the Kalman filter prediction step in order to detect the new locations of monitored pedestrians and potential new pedestrians.

D. Fusion-based Layered-based Tracking

The initial segmentation results, human locations, are used to initialize dynamic tracking models for human positions. We define \( N \) Kalman filter (KF) model for \( N \) pedestrians as described in equation (1).

\[
\begin{align*}
\mathbf{x}_k &= A\mathbf{x}_{k-1} + Bu_{k-1} + w_{k-1} \\
\mathbf{z}_k &= H\mathbf{x}_k + v_k \\
p(w) &= N(0, Q) \\
p(v) &= N(0, R)
\end{align*}
\] (1)

where \( \mathbf{x} \) is the state, \( \mathbf{z} \) is the measurement, \( w \) and \( v \) are respectively process noise and measurement noise, \( Q \) and \( R \) are respectively process noise covariance and measurement noise covariance.

The states of each KF model include the horizontal position \( \text{pos}_x \) and horizontal speed \( v_x \) for each pedestrian, i.e., \( \mathbf{x} = [\text{pos}_x \ v_x] \). We set \( B = [0; 0] \), \( H = [1 \ 0] \), \( A = [1 \ 1; 0 \ 1] \) since \( \text{pos}_x[k] = \text{pos}_x[k-1] + v_x[k-1] \) and \( v_x[k] = v_x[k-1] \).

Kalman filtering involves the step of estimation and update. For estimation step, we estimate horizontal locations and speeds for humans, and the error covariance in the next frame based on historical data as shown in equation (2).

\[
\begin{align*}
\hat{x}_k &= A\hat{x}_{k-1} + Bu_{k-1} \\
\hat{P}_k &= AP_{k-1}A^T + Q
\end{align*}
\] (2)

For update step, the measured horizontal locations and speeds of previously identified pedestrians are used to update the gain, measurement, and error covariance of KF models as shown in equation (3).

\[
\begin{align*}
K_k &= P_k^{-1}H^T(HP_k^{-1}H^T + R)^{-1} \\
\hat{x}_k &= \hat{x}_k + K_k(z_k - H\hat{x}_k) \\
P_k &= (I - K_kH)P_k^{-1}
\end{align*}
\] (3)

Different from traditional Kalman filtering where predicted results from equation (2) are only intermediate results, we actively take advantages of the predicted horizontal location and speed of pedestrians. The information is sent to a fusion-based tracking block to improve its reliability. As shown in the flowchart in figure (6)(a), fusion-based tracking block receives information from initial segmentation and the Kalman filter prediction step in order to detect the new locations of monitored pedestrians and potential new pedestrians.

Figure (6)(b) introduces how fusion-based tracking block fuses the information from initial segmentation and KF prediction to identify the new location of pedestrians. The fusion-based tracking block accepts the following information:

1) Initial segmentation for the current frame
2) Vertical projection-curves for the current frame based on three continuous images
3) Predicted horizontal locations and speed of pedestrians from KF prediction

![Fig. 5. The changing trend of defined vertical projection-curves and the results of segmentation and tracking before and after two people intersected.](image)

![Fig. 6. (a) Algorithm flowchart. (b) Flowchart for fusion-based tracking.](image)

551
The block also stores the historical information of vertical projection curves, body-trunk templates and other related pedestrian information.

Based on KF prediction and horizontal locations of triangle spikes in previous vertical projection curves, the possible new horizontal regions for all pedestrians can be predicted. We then check whether there exist independent sharp triangle spikes at the predicted regions in order to determine potential intersecting cases or new independent waves.

If all triangle peaks show up at estimated regions and there are not merging peaks, we can directly use the initial segmentations and apply body-trunk-based matching to associate the detected pedestrians to the stored pedestrian information. If there are new additional triangle peaks that do not correspond to any previous pedestrians, the corresponding initial segmentation provides the location of new pedestrians. For both cases which do not involve occlusion, initial segmentation is reliable which will be sent to update dynamic KF models and to update the stored information of body trunk template and other related pedestrian information.

If triangle spikes fail to show up at estimated regions, they might merge with other peaks or disappear. Corresponding pedestrians might be occluded or walk away. When two original independent triangle spikes start to merge, it shows that two pedestrians move closer and intersection (occlusion) starts. Initial segmentation only captures merged regions containing two interested objects. However, figure (4) shows that the horizontal locations for two pedestrians are at transitional peaks of the vertical projections as mentioned in section II-B. Thus, we search for detail trunk locations by comparing previous human body trunk templates with corresponding image regions at all transition peaks within merged waves. After determining the locations of body trunks, we reuse the size information from previous frames to describe human regions during intersect (occlusion). Thus our proposed methods can reliably detect the locations of both pedestrians in most intersection (occlusion) frames. The method might have problems when the horizontal locations of two people’s heads are completely the same, i.e., when one person is completely occluded by another. However, for most occlusion situations, the horizontal locations for two pedestrians are at different transitional peaks during the whole merging process though the images of two people might merge in several continuous frames. It is impossible for one person to be completely occluded by another in several frames unless they are static or both walk together. In that case, they are treated as one group.

For occlusion cases, we do not update KF dynamic models and pedestrian trunk templates as shown in figure (6). Segmentation reliability might degrade because of matching ambiguity when searching for head area among wave peaks. Human body trunk area might consist of texture information from another person since interested pedestrians do not completely show up. We use the same matching templates during occlusion to avoid the impact of incorrect measurement data on the performance of Kalman filter based dynamic models. For pedestrians who disappeared, body-trunk templates and tracking models are updated accordingly.

E. Results

Figure (5) shows the results of segmentation and tracking based on the corresponding projection-curves. Specifically, figure (4) shows the human detection details before and after two people intersect. The horizontal locations of people’s head are exactly at the transitional peaks of the projection-curves, and the horizontal locations of merged regions are accurately determined by peak boundaries of projection-curves. The algorithm accurately captures two people during the intersection.

To demonstrate the performance in a long sequence, we present detail locations of each detected person in specific frame by plotting a square black dots at the corresponding head locations in image frames. One person’s movement in a whole sequence is then represented by a dotted line in one fixed background image which is the average of the first and the last images and contains the initial and the final appearance of pedestrians. Figure (7) shows the integrated tracking results for the whole sequences. For each sequence, the number of dotted lines should be equal to the number of frames in which pedestrians show up. Besides, the dotted line describes the moving path of the pedestrians, which should be smooth because of smoothness of human movement. Figure (7) does provide reliable and correct tracking results.

Figure (8) and (9) and figure (10) and (11) respectively show the tracking results for two video sequences when pedestrians change their poses. For sequence in figure (8)(9), one pedestrian gradually changed from his standing pose to bowing pose. For sequence in figure (10)(11), the pedestrian first walked to the right, then turned 90 degrees and walked to the back. Figure (12)(13) shows the tracking results when one pedestrian first walked backward and then forward, while
Fig. 9. Tracking results of two people for whole sequences in figure (8). Left: Pedestrian 1. Right: Pedestrian 2.

Fig. 10. The segmentation and tracking results before and after two people intersected. One person turned 90 degree and changed the direction of walking.

Fig. 11. Tracking results of two people for whole sequences in figure (10). Left: Pedestrian 1. Right: Pedestrian 2.

Fig. 12. The segmentation and tracking results before and after two people intersected. One person first walked backward and then forward.

Fig. 13. Tracking results of two people for whole sequences in figure (12). Left: Pedestrian 1. Right: Pedestrian 2.

F. Performance Evaluation

We tested 26 video sequences where there were two or more people walking toward different directions with different poses. Our evaluation standard is whether the algorithm can successfully and accurately capture the same person all the time. When people are lost during the tracking process, and are later detected as a new person, the total number of detected people will be larger than real human number. If that happens, or the motion path for that person is incorrect, especially when the person is lost after intersection, we claim “failed” for the sequence. We claim “success” when motion path is detected accurately with some segmentation errors.

Among 26 test sequences, 3 test sequence results are “failed” (around 12%) and there are 23 “succeed” test results (around 88%).

To compare the performance of our proposed algorithms with conventional methods, we applied conventional segmentation based background subtraction and mean-shift tracking algorithms to track pedestrians’ behavior in the same sets of occluded sequences. The rate of “success” is only 40%. Besides, the results based on conventional methods take advantage of the color information while our algorithm uses only grey level images for the same set of sequences.

The current detection results show that our algorithms have the capability to track humans with changing walking-poses, different walking directions, and to deal with segmentation with partial occlusions and scale variations of targets. Humans are accurately tracked even if merging and intersecting happens. The proposed algorithm has significantly improved the detection reliability from 40% to 88%.

III. DISCUSSION: THE FEATURES & ADVANTAGES OF PROPOSED ALGORITHMS

Our proposed layered-based and fusion-based algorithms have the following features and advantages.

A. The Advantage of Layered-based Processing based on Vertical Projection-Curves

Our segmentation algorithm follows the layered-based principle. Taking advantages of defined vertical projections, we can first separate an image into several vertical stripes and identify the stripes containing interested pedestrians. Then we search for the vertical locations of potential pedestrians. Our “projection-based, horizontal first, vertical next” segmentation changes the original two-dimensional search of pedestrians within a whole image into two one-dimensional search, which significantly improves segmentation reliability and decreases the computational load.

The two 1D searching scheme helps to simplify the dynamic tracking model. Our dynamic models only track the horizontal location and speed. In contrast, traditional complicated tracking models track shape, contour, and location information, at least track both horizontal and vertical locations of interested
objects because of limited reliability and accuracy of vertical detection during occlusion situations. Our simple dynamic models are reliable enough to provide good estimation for future horizontal locations of interested pedestrians. The detection noises are still Gaussian and Kalman filter works pretty well in our cases. The projection-curve-based algorithm does not depend on human features, such as faces, skins, etc., ellipse shape, etc., thus having more general applications than detection algorithms relying on these special features.

1) The Limitation of Traditional Algorithms for Initial Segmentation: In comparison, the traditional methods for initial segmentation, such as reference-image-based segmentation, feature-based segmentation, and geometric-model-based segmentation, are still very limited in their reliability and computational speed.

Reference-image-based segmentation is to detect moving objects based on the differential images between video frames and reference images. If cameras are fixed (surveillance situations), a reference can be background images estimated through statistical model [3] or acquired when no foreground shows up [12] [18]. Foregrounds are detected from differential images through k-means clustering based on motion information [20] [21]. Typically motion information estimation is noisy, especially at occlusion boundaries. Human body is not rigid and their motion is not uniform. Furthermore, segmentation accuracy and reliability determined by the quality of differential images would be affected by changing lighting and shading situations, shape-changing non-rigid background objects such as trees branches, etc.

Feature-based segmentation detects humans by capturing their distinctive features, such as human faces or heads [13] or skin color [14] [17], the rigidity of KLT interested corner feature points in head/header areas, depth information, deviation from given human template, etc. These features might not exist for duration occlusion, which limits the application. Finding a general template for pedestrians with various poses is difficult.

Geometric-model-based segmentation generates a specific human geometric model that describes human beings, including three dimension human models /high DOF articulated models for bodies and/or limbs [15] [10], two dimension appearance-based model (blob-based / pixel-based texture model [11], a texture-free shape model based only on Mahalanobis distance map [16], statistical for- and background models, illumination based model) [20], etc. These models build the connection between different poses and their corresponding images. Complicated human configuration model is usually needed for pose analysis, which usually involves computational cost and might not work very well for occlusion situations [19].

Compared with traditional segmentation/tracking algorithms purely based on template matching, the proposed algorithm is simple, accurate and effective.

B. The Advantage of Fusion-based Principle

In order to enhance tracking accuracy, our fusion-based scheme takes advantages of combined information, including, initial horizontal segmentation, defined vertical projection-curves, estimated human locations, and body trunk templates.

1) Tracking based on Fusion of Initial Segmentation and Dynamic Tracking: Traditionally, pedestrian detection involves initial detection based on human’s features for one frame and tracking for several frames afterward. Many algorithms cannot afford implementing segmentation for every frame because of its heavy computational load. The computational load of our initial segmentation is very light, thus we can afford to implement initial horizontal segmentation and to update body-trunk templates for every frame which allows for more adaptability in situations with rapid pose changes.

Then we take advantage of the similarity of pedestrian trunk template to associate the human position information in different frames. The performance of our initial segmentation is very high and provides satisfying segmentation results for non-occluded humans, which significantly decreases the matching ambiguity during tracking. The combination of segmentation and dynamic tracking does not need long initialization time, does not depend heavily on tracking, and can quickly respond if people dramatically change their poses. The final detection system is simple but effective in reducing the segmentation/tracking ambiguities when human intersects and occurs.

2) Tracking based on Fusion of Vertical Projection-Curves, Historical Information and Tracking Features: The special defined vertical projection-curves are not only useful for searching for horizontal locations of potential pedestrians during initial segmentation, but also help to decrease the searching ranges when matching pedestrians among different frames. We only need to compare human trunk template from previous frame with very few candidate regions corresponding to the triangular peaks of projection-curves within estimated location regions. Even for merging cases, the search ranges for candidate pedestrians can be limited to transitional peaks within the merged projection-curves. Thus, the fusion-based tracking is very useful to identify individual pedestrians with various poses, including the cases when people bow (which is very rare in most monitoring situation) as shown in Figure (8) (9).

The predicted information from KF models helps to identify the potential merging cases. When the reliability is low, tracking processes stop updating the template and location variables in dynamic models with new measurement data, which prevents the ambiguity and failure during occlusion to pollute the dynamic model and improve estimation performance and robustness for unexpected situations. The mechanisms to update tracking information based on the reliability of segmentation and tracking improves system robustness when dealing with serious tracking failure.

3) The Limitation of Traditional Algorithms for Human Tracking: In contrast, complicated math models are involved in traditional tracking schemes. Mean Shift Iteration algorithm [24], a nonparametric estimator of density, tracks non-
rigid objects based on visual features whose statistical distributions characterize the object of interest. The technique works for objects with different color/texture patterns, and are robust to partial occlusions, clutter, rotation in depth, and changes in camera positions. However, mean shift iterations are needed to find target candidates most similar to a given target tracking models with specific similarity metric, say, Bhattacharyya coefficient. Geometric model based algorithm tracks human geometric models parameters, for example, human body contour curves [19], human body regions (blobs) [25], pose and location parameters of bodies and hands, etc. Such tracking models are usually not uni-modal, involve transformation, and usually are not Gaussian [19]. The complicated algorithms need a few frames’ initialization time before tracking performance gradually stabilizes. Particle filtering [22] and related algorithms require no model and support multi-modal tracking, and thus they can deal with nonlinear and non-Gaussian problems for stochastic dynamic estimation at the cost of higher computational load. [23] formulates human segmentation as a Bayesian MAP estimation problem whose solution is pursued by Markov chain Monte Carlo approach. Because of heavy computational load, most tracking techniques work better with larger and slower objects and are limited in their ability to track small fast objects.

IV. CONCLUSION

This paper proposes a layered-based and fusion-based human detection and tracking system that can handle human intersection and conclusion in complicated environment. The proposed algorithm applies layered-based initial segmentation for every frame based on “projection-based, horizontal first, vertical next” segmentation scheme. In handing occlusion, our mechanism chooses potential head locations at local projection peaks within predicted regions. Kalman-filter-based dynamic models are used to predict pedestrians’ horizontal location and speed. The fusion-based tracking process identifies and keeps track of detected pedestrians through fusing the information from the initial segmentation, projection-curves, prediction results from KF filters, and historical information of pedestrians’ body-trunk templates. The final detection results are used to update human locations in Kalman filters and human matching templates. Compared with other human detection methods, our layered-based fusion-based algorithm is very simple and avoids complicated tracking models while it is very effective in detecting and tracking the movements of pedestrians through partially occluded situations.

REFERENCES