OCCLUSION HANDLING FOR PEDESTRIAN TRACKING USING PARTIAL OBJECT TEMPLATE-BASED COMPONENT PARTICLE FILTER

Daw-Tung Lin. Department of Computer Science and Information Engineering, National Taipei University, Taiwan.

Yen-Hsiang Chang. Department of Computer Science and Information Engineering, National Taipei University, Taiwan.

ABSTRACT
Pedestrian tracking plays a crucial role in security and intelligent video surveillance. Occlusion handling is a challenging concern in tracking multiple people. Adaptive, advanced solutions are required to accurately track pedestrians for video surveillance purposes. This paper presents a novel method of tracking multiple people in occlusion conditions. In this study, we developed a combined component-based human-shaped template and a particle filter, resulting in improved object occlusion handling. The proposed system is capable of tracking specific people in real-time, as well as handling the occlusion of multiple objects. We predicted occlusions using the Kalman filter, and tracked objects continuously using our component-shape-template particle filter. The experimental results show that our algorithm is feasible and stable. In tests, the proposed tracking algorithm achieved an accuracy of up to 99.7%. The proposed approach outperformed that of comparable methods in analyzing all test video datasets in this study. Our low false-negative rate demonstrates that the proposed tracking method is robust and offers superior occlusion handling.

KEYWORDS
Pedestrian tracking, occlusion handling, video surveillance, template-based component matching, particle filter.
1. INTRODUCTION

In recent years, traditional video monitoring systems have been overwhelmingly replaced by intelligent video monitoring systems. Intelligent video surveillance plays a crucial role in security. Applied to public space design, visual surveillance, and the construction of intelligent environments, pedestrian detection and tracking are a major challenge for computer vision system architecture (Enzweiler & Gavrila, 2009; Zhan et al., 2008). The two major approaches to object tracking are appearance-based (Erdem et al., 2003; Luo & Eleftheriadis, 2003; Van Beeck et al., 2011) and motion-based (Kim & Hwang, 2002; Lin & Huang, 2011) algorithms. Rosales and Sclaroff used the extended Kalman filter to estimate the 3D trajectory of an object in 2D motion (Paek et al., 2007). Particle filter is one of the appearance-based methods and has been frequently used in tracking. Ma et al. (2007) presented a spatial-color object model and developed an efficient, particle-filter-based visual tracking algorithm. Paliao and Batista (2008) represented objects using covariance matrices and applied a particle filter to perform object tracking. Li et al. (2008) combined color distribution histograms with a particle filter, and examined target shapes as a necessary factor in creating target models. Huang et al. (2008) performed a shape analysis on foreground blobs and foreground detection likelihood outputs using particle-filter-based object tracking. Shan et al. (2007) used particle filter and mean shift algorithms to track hand gestures. Reuter and Dietmayer (2011) used a sequential Monte Carlo multitarget Bayes filter based on random finite set theory to perform pedestrian tracking. Hu et al. (2012) modified the state transition model of particle filters using an optical flow algorithm, and proposed a new tracker. Guan et al. (2013) modeled particle filter dynamics as a second-order autoregressive process and improved the observation model using a color histogram likelihood measure. Choi and Yoo (2013) devised a particle filter framework that used object size estimation and tracking failure detection to improve tracking accuracy and robustness.

Occlusion renders the detection and tracking of people difficult to perform. Adaptive, advanced solutions are required to track pedestrian behavior for video surveillance. To overcome the occlusion problem, several approaches have been developed. Enzweiler et al. (2010) proposed a novel mixture-of-experts framework for pedestrian classification and partial occlusion handling. Ess et al. (2009) combined classical geometric world mapping with multiperson detection and tracking, which estimates camera position, stereo depth, object detections, and trajectories to improve tracking accuracy. Li et al. (2008) improved the traditional mean shift tracking algorithm by using occlusion layers to represent the pedestrian occlusion relationship and adjusting the states of related pedestrians to eliminate the effects of occlusion during the tracking process. Ablavsky and Sclaroff (2011) formulated a layered graphic model for tracking partially occluded objects, defining an occluder-centric representation as a first-order Markov process on activity zones with respect to the occlusion mask of the relocatable object. Singh et al. (2008) presented a two-stage multi-object tracking approach to robustly track pedestrians in occlusion scenarios by generating high-confidence partial track segments and associating the tracklets in a global optimization framework. Corvee and Bremond (2010) presented a hierarchical tree of histogram of oriented gradients (HOG), and coupled it with independently trained body-part detectors to enhance detection performance. Chong et al. (2012) combined multiple color and edge feature cues in a particle filler framework to resolve object tracking problems when occluded by other human bodies and structures. Airouche et al. (2012) proposed the Dezert-Smarandache (DSm) theoretical
framework, which uses location, color, and thermal feature measurements to track pedestrians in cluttered spaces. Kratz and Nishino (2012) employed a collection of hidden Markov models, trained on local spatiotemporal motion patterns, to represent crowd motion; they demonstrated that it is possible to track people in crowded spaces using a prior-movement-prediction model. Ge et al. (2012) performed automatic crowd analyses in complex situations using a generalized and symmetric Hausdorff distance defined with respect to pairwise proximity and velocity. Führ and Jung (2013) presented a new approach combining vertical standing patch matching and a weighted pedestrian detection vector; this approach is able to address occlusions and video sequences with strong appearance variations.

The objective of this study was to develop a robust tracking system to use in occlusion conditions. The Kalman filter was applied to evaluate velocity and predict the positions of objects. However, the Kalman filter does not work well on objects moving in a nonlinear fashion. We therefore adopted a particle filter method to track nonlinear movement. A robust occlusion detection algorithm was developed in this study, to enable tracking when two or more objects are partially occluded. We further modified our particle filter algorithm, and we propose a template-based component-matching method to robustly and accurately handle various occlusion situations during the tracking process.

This paper is organized as follows. Section 2 illustrates the architecture of the proposed tracking system. Section 3 elucidates the main technique used by the proposed occlusion handling method, which consists of occlusion detection and object template-based component matching using a particle filter. In Section 4, the experimental dataset is described and the simulation results are presented. Finally, the conclusion is discussed in Section 5.

2. OBJECT TRACKING SYSTEM ARCHITECTURE

![Figure 1. The overall block diagram of the proposed tracking system.](image)

Occlusion is a challenging problem in object tracking. It is difficult to identify objects using occlusion, because foreground objects cannot be completely extracted. In this paper, we propose a method to resolve the occlusion problem. Figure 1 shows a block diagram of the proposed tracking system, which contains two subsystems: video object correspondence and an occlusion handling algorithm.
OCCLUSION HANDLING FOR PEDESTRIAN TRACKING USING PARTIAL OBJECT TEMPLATE-BASED COMPONENT PARTICLE FILTER

After identifying moving objects, the method tracks moving objects in subsequent frames. Tracking is a process that continuously searches for the best object matches between the current and previous frames. The features of particular objects are recorded to enable more effective subsequent prediction of object locations. Let $VO_i^n$ denote a new object $i$ extracted from current frame $n$. To match $VO_i^n$ with an object $VO_j^{n-1}$ appearing in previous frame $n-1$, a feature-matching function $FM(VO_i^n, VO_j^{n-1})$ is calculated to estimate the similarity between two objects $VO_i^n$ and $VO_j^{n-1}$. Color histogram and object size are utilized as object matching features. The histogram information and physical properties are related as follows (Equation (1)):

$$FM(VO_i^n, VO_j^{n-1}) = \alpha \cdot PHist(VO_i^n, VO_j^{n-1}) + \beta \cdot PCnt(VO_i^n, VO_j^{n-1}).$$

where $\alpha$ and $\beta$ are the linear combinational weights of features $PHist(VO_i^n, VO_j^{n-1})$ and $PCnt(VO_i^n, VO_j^{n-1})$, respectively, and $\alpha + \beta = 1$. $PHist(VO_i^n, VO_j^{n-1})$ and $PCnt(VO_i^n, VO_j^{n-1})$ denote the similarity probability of the histograms and object sizes of $VO_i^n$ and $VO_j^{n-1}$, respectively. The similarity measure of the histogram $PHist(VO_i^n, VO_j^{n-1})$ is expressed as:

$$PHist(VO_i^n, VO_j^{n-1}) = \frac{1}{K} \sum_{z=1}^{K} \frac{C_z}{A_z + B_z - C_z},$$

where $K$ is the number of histogram bins, $C_z$ is the minimum value of two bins (i.e. $C_z = \min(A_z, B_z)$), $A_z$ denotes the $z$th bin of histogram of $VO_i^n$ and $B_z$ denotes the $z$th bin of histogram of $VO_j^{n-1}$, and $1 \leq z \leq K$.

The similarity measure of the object size $PCnt(VO_i^n, VO_j^{n-1})$ is as follows.

$$PCnt(VO_i^n, VO_j^{n-1}) = 1 - \left| \frac{u_i - u_j}{u_i + u_j} \right|,$$

where $u_i$ and $u_j$ represent the number of pixels of $VO_i^n$ and $VO_j^{n-1}$, respectively. To improve the matching accuracy, the object movement features are adapted, such as velocity and position as illustrated in the next section.
3. OBJECT OCCLUSION HANDLING

3.1 Occlusion Detection

![Figure 2](image)

Figure 2. The process of occlusion detection: (a) Frame n, (b) Frame n+1 with occlusion detection, (c) result of occlusion detection (two detected objects are drawn in red and blue line, respectively).

In this section, we describe how objects can be identified when two or more objects are partially occluded. When an object is occluded, the method conserves the features of that object and updates its prediction of the object’s future position. After updating the prediction of the object’s future position, the method estimates the position of the occluded object and determines whether the object is hidden by other objects. Figure 2 shows a diagram of the proposed occlusion detection algorithm. Generally, the predicted position of an occluded object is usually located in the vicinity of what appears to be a new, previously undetected object, but is actually already-observed objects in a condition of occlusion. It is possible to evaluate how many objects from the previous frame remain in the vicinity of new object in the current frame.

As illustrated in Fig. 2(a), two objects $VO_0^i$ and $VO_1^i$ appear in frame $n$ and are moving in different directions. Next, as shown in Fig. 2(b), $VO_0^i$ and $VO_1^i$ occlude with each other and form a new object $VO_{1+i}^i$ as seen in frame $n + 1$; at this point, objects $VO_0^i$ and $VO_1^i$ are no longer detectable in frame $n + 1$. However, we can observe that the predicted centroid positions of $VO_0^i$ and $VO_1^i$ are in the area of $VO_{1+i}^i$. Fig. 2(c) shows the tracking results and combines a rectangle mask and the predicted position to identify occluded objects. The proposed occlusion detection algorithm is summarized as follows:

Step 1: Predict the position of motion object $VO_{i+1}^i$ in frame $n - 1$ using the Kalman filter, record the object’s feature and update its predicted position.

Step 2: Check every object $VO_i^j$ to determine whether more than one predicted centroid of $VO_{i+1}^j$ is located in the area of $VO_i^j$.

Step 3: Apply the similarity matching function (Equation (1)) and analyze the occluded video object $VO_{i+1}^j$. 
3.2 Object Template-based Component Matching with Particle Filter

Particle filters are a popular object tracking method and can reliably predict the positions of nonlinear moving objects. To further improve the tracking accuracy of the method in occlusion conditions, we combined a method to analyze color features and a particle filter.
When the particle filter algorithm updates the predicted positions of objects based on new measurements, the method uses the Bhattacharyya distance $D$ to evaluate the similarity of hue value histograms using the object’s HSV color model between each particle and object $VO_{i-1}$.

When the value of the Bhattacharyya parameter is equal to 1, the two color histograms are equal. After obtaining the distance distribution, the measurement likelihood function can be selected as follows:

$$p(Z_i | X^t_i) \propto N(D;0,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-D^2}{2\sigma^2}\right)$$

(4)

When this equation yields a greater value, it means that the color correlation histograms of each particle and $VO_{i-1}$ are more similar. After estimating the posterior mean state, the result presents the position of $VO_{i-1}$.

Traditional particle filters use rectangle or ellipse masks to define the sampling region for each particle. These rectangles or ellipses are of fixed shape and fixed size. We revised this sampling technique to achieve a more reliable tracking accuracy. Figure 3 illustrates the traditional particle filter approach, which spreads particles in the rectangle mask shown in Fig. 3(b) and uses mean movements to estimate tracking locations, and thus yields false predictions when confronted with occlusion. To resolve this problem, we present a novel procedure that utilizes a component particle filter (Liu, 2008). First, a rectangle mask is divided into four components as shown in Fig. 4(b). If the object being observed is partially occluded, it is still possible to accurately track the object using the other components. However, because people are irregularly shaped, rectangle masks are not that suitable for tracking people. To further improve tracking accuracy, our proposed system uses the shape of objects to create a template-based component mask and revise the component particle filter where particles only selected in the specific object template regions. The template-based component masks are based on the shape of $VO_{i-1}$. An example of this is shown in Fig. 5(b), where the component mask of $VO_{i-1}$ is obtained from the shape of $VO_{i-1}$. This mask is then divided into four components. Each template of object $VO_{i-1}$ is unique, so the particle filter’s matching results are not only based on the similarity measure of the color histograms and object sizes, but also on shape similarity. As illustrated in Fig. 5(c), although one of the masks failed in tracking because of occlusion, the other masks successfully tracked objects with the corresponding components, because $VO_{i-1}$ was only partially occluded, and un-occluded regions of $VO_{i-1}$ could still be detected.

4. EXPERIMENTAL RESULTS

The proposed template-based component particle filter tracking system was implemented in Visual C++ 7.0 using a Windows platform. The developed system was tested in experiments on four video clips: two test videos recorded on our university campus and two videos obtained from the PETS2006 evaluation dataset. The ground truth of pedestrian motion was manually generated using these four video clips. We performed multiple object tracking on an Intel Core 2 Duo 2.4 GHz processor with 2 GB Ram and a video resolution of 320 x 240 which performed at a speed of 10 frames per second. Figs. 6 and 7 display some of the tracking results. The red and blue rectangles represent tracking locations in non-occlusion and
occlusion conditions, respectively. The same ID in different frames indicates that pedestrians were successfully tracked through each consecutive video sequence.

We compared the results of our pedestrian tracking system with results obtained using two well-known algorithms, the mean shift and traditional particle filter algorithms, and our previous work, four-part particle matching (Liu, 2008). Two types of evaluation indices were adopted: frame-based statistics and object-based counting. Frame-based evaluation consisted of three measurements: correct tracking (whether the correct number of objects were tracked in each frame), missing tracking (the number of object tracks lost), and false tracking (the number of objects accidentally identified as new objects or false positives). The object-based indices we adopted to evaluate tracking performance (Senior et al., 2006) were the false positive rate $fp$ and the false negative rate $fn$, as ratios of object tracking numbers.

Table 1 summarizes the performance of these tracking methods based on these evaluation indices. The proposed tracking algorithm achieved an accuracy of up to 99.7% when tested on videos recorded on our university campus (Video 1 and Video 2). The proposed approach outperformed all other tracking methods on all test videos. The high missing tracking numbers of all methods in the analysis of the PETS2006 S-T1-G test video (Table 4) was because certain pedestrians in the video were too small to be detected, for example, the person with ID label 8 shown walking from right to left in Fig. 7(a). In the same video, the person with ID label 0 is occluded (shown in Fig. 7(b)), and Fig. 7(c) shows that this person was still tracked successfully. Our low false-negative rate reveals that the proposed tracking method is robust in occlusion handling.

![Figure 6](image-url)
Figure 7. The results of pedestrian tracking tested on PETS2006 S-T1-G.

Table 1. Performance comparison of tracking accuracies obtained by different methods measured on four video clips.

<table>
<thead>
<tr>
<th>Video</th>
<th>Tracking Algorithm</th>
<th>Frame Measure</th>
<th>Object Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Correct Tracking</td>
<td>Miss Tracking</td>
</tr>
<tr>
<td>Video 1</td>
<td>Mean Shift</td>
<td>3571</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>3587</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>Four-parts Matching</td>
<td>3588</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Our Method</td>
<td>3592</td>
<td>7</td>
</tr>
<tr>
<td>Video 2</td>
<td>Mean Shift</td>
<td>3705</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>3759</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Four-parts Matching</td>
<td>3755</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>Our Method</td>
<td>3785</td>
<td>34</td>
</tr>
<tr>
<td>PETS 2006 S5-T1-G</td>
<td>Mean Shift</td>
<td>2627</td>
<td>1308</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>2500</td>
<td>1435</td>
</tr>
<tr>
<td></td>
<td>Four-parts Matching</td>
<td>2495</td>
<td>1440</td>
</tr>
<tr>
<td></td>
<td>Our Method</td>
<td>2820</td>
<td>1115</td>
</tr>
<tr>
<td>PETS 2006 S3-T7-A</td>
<td>Mean Shift</td>
<td>1809</td>
<td>531</td>
</tr>
<tr>
<td></td>
<td>Particle Filter</td>
<td>1781</td>
<td>559</td>
</tr>
<tr>
<td></td>
<td>Four-parts Matching</td>
<td>1836</td>
<td>504</td>
</tr>
<tr>
<td></td>
<td>Our Method</td>
<td>2192</td>
<td>148</td>
</tr>
</tbody>
</table>

5. CONCLUSION

We developed a tracking algorithm combining a component-based human shape template and a particle filter to improve object occlusion handling. The proposed system is capable of tracking specific people in real-time, as well as handling the occlusion of multiple objects. We predicted occlusions using the Kalman filter and tracked objects continuously using our component-shape-template particle filter. The experimental results show that our algorithm is feasible and stable. The proposed tracking algorithm achieved an accuracy of up to 99.7% when tested on a video recorded on our university campus. The proposed approach outperformed two well-known algorithms as well as our previous work on all test video datasets. Our low false-negative rate revealed that the proposed tracking method is robust and superior in occlusion handling.
ACKNOWLEDGEMENT

This work is partially supported by the National Science Council under project funding contract Nos. NSC 99-2221-E-305-012-MY2 and NSC 101-2221-E-305-011-MY2.

REFERENCES


