AN ADAPTIVE APPROACH FOR OVERLAPPING PEOPLE TRACKING BASED ON FOREGROUND SILHOUETTES

Hsin-Ho Yeh[1], *Jiun-Yu Chen*[1], *Chun-Rong Huang*[1] *and Chu-Song Chen*[1,2]

[1]Institute of Information Science, Academia Sinica, Taipei, Taiwan.
[2]Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan.
{hhyeh, jychen, nckuos, song}@iis.sinica.edu.tw

ABSTRACT

We propose **B**inary/**A**ppearance **Tracker** which consists of background subtraction, silhouette similarity and particle filter to infer pedestrians' locations under different occlusion situations with a single camera. During the period of occlusions, binary and color silhouettes are adaptively used to effectively measure the similarity between the observation and the possible combinations of silhouettes. Thus, the occluded pedestrians' locations can be simply located by the most possible combination of silhouettes. The experimental results show that the proposed BATracker can track people successfully even though she/he is fully occluded.

Index Terms— Particle filter, Multiple people tracking

1. INTRODUCTION

This paper presents a pedestrian tracking approach, particularly for occlusion situations. Pedestrian tracking is important for vision-based surveillance applications. For the past decade, general object tracking methods have been proposed [1]. Many methods [2][3][4][5] used an appearance model to track objects, but the appearance model is easily affected by illumination changes, which cause the tracker to drift. Besides appearance, shape is another popular information used for tracking [6][7], but it suffers from the intrinsic variance [3] caused by the pose changes of moving people.

Another difficulty in pedestrian tracking is occlusion. To handle the occlusion problem, [2] assumes a constant motion model to predict the possible locations of the occluded objects in the next frame. The constant motion model cannot describe human motions well in real environments, particular in crowded scenes. To improve pedestrian tracking under occlusions, particle filter (PF) with inertia motion model is used to predict pedestrians' trajectories [4][5]. A recent novel pedestrian-tracker [4] relies on the confidence of the pedestrian detector, online classification and PF to infer pedestrians' locations, but it requires prior knowledge that has to be learned for a period of time before it starts detection. In [5], the ellipse template, appearance model and PF are cooperated to deduce the occlusion relationships between multiple people. The ellipse shape, however, cannot model the human shape precisely due to intrinsic variances.

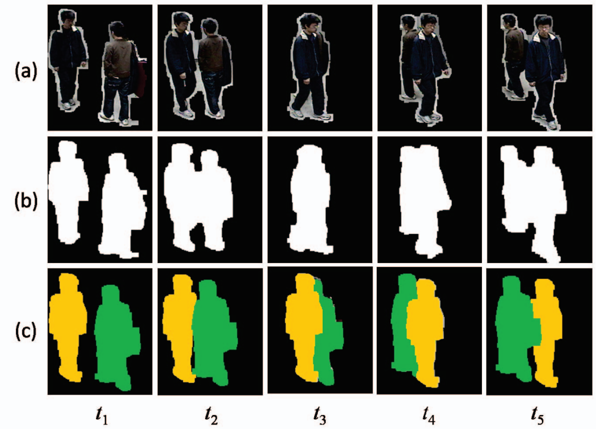

Fig. 1. The columns show some examples of different occlusion situations, where t_1 is the case of non-occlusion, t_2 and t_5 show the case of slight occlusions, and t_3 and t_4 present the case of heavy occlusions, respectively. Rows (a) and (b) show the colored silhouettes and binary silhouettes, respectively. Note that, for the slight occlusion case (t_2 and t_5), shape information (row (b)) is sufficient enough to infer the pedestrians' locations (as shown in row (c)). However, for the heavy-occlusion case (t_3 and t_4), combination of both shape and appearance information (row (a)) will be necessary.

In this paper, we propose a silhouette-based tracking method, called **B**inary/**A**ppearance **Tracker** (**BATracker**), where binary and appearance are referred to the shape and color-appearance information, respectively. We use the foreground blobs (or called silhouettes) to infer pedestrians' locations by PF under different occlusion situations, so that the multiple occluded-pedestrians tracking problem can be handled with a single camera. The silhouette-based approach is more suitable for human shapes than the ellipse-based model [5], where the silhouettes are obtained from background subtraction [8] in our work. The used information is adaptive to the following situations:

(1) Non-occlusion: The tracking problem is reduced to a simple blob association problem.

(2) Slight occlusion: We only use the shape information (i.e. binary silhouette) for tracking, since the overlapped shapes can be informative for pedestrian location inference. This is motivated by [6].

(3) Heavy occlusion: Using only the shape information is not sufficient. We additionally exploit the appearance information, that is, the colored silhouette, for tracking.

Fig. 1 gives an illustration of the above situations. In this paper, we propose a PF-based approach that can adaptively deal with the above situations by switching the type of information used in our approach. An efficient method is thus conducted to handle pedestrians' occlusions under different situations for tracking.

2. OVERVIEW

As mentioned above, BATracker tracks occluded pedestrians via inference in different situations. For a coming video frame, foreground pixels obtained from background subtraction are grouped into foreground silhouettes (FS) according to the connectivity after morphological operations. As no occlusions, an FS in the current frame is simply associated to its nearest FS (in terms of the distance between their centroids) for tracking. When occlusions occur, we consider the probability distribution of hypothesized silhouettes, and try to find the maximum likelihood hypothesis by using PF. For example, as shown in Fig. 1, the silhouettes in five consecutive frames t_1, t_2, t_3, t_4, t_5 are shown in each column. Two silhouettes exist in frame t_1, but they are occluded in frames t_2, t_3, t_4, t_5. We name the silhouettes in frame t_1 as initial silhouettes $S_{t_1} = \{s_{t_1}^1, s_{t_1}^2\}$ and the silhouettes in frames t_2, t_3, t_4, t_5 as O_{t_2}, O_{t_3}, O_{t_4}, O_{t_5}, respectively. During the occlusion period, we want to infer the people locations from S_{t_1} and O_{t_n}, where $t_n \in \{t_2, t_3, t_4, t_5\}$. Note that, although the above example is illustrated for two-people occlusion case, the symbols and terms can be easily generalized for multi-people occlusion cases.

3. ADAPTIVE PF BASED ON LIKELIHOOD SWITCHING

In this work, we use an ordered sequence as a layered representation of the overlapped silhouettes. A silhouette sequence $\sigma = \left(s^1, s^2, ..., s^M\right)$ is employed to define the situation that these silhouettes are overlapping to each other and s^k is above s^l if $k < l$. Given a silhouette s^k, the operation $s^k + v$ stands for shifting the silhouette by a 2D vector $v = (v_x, v_y)$. The symbol H_σ defines the region covered by the overlapped silhouettes. For example, consider the initial silhouettes $\{Y, G\}$ shown in Fig. 2(a). Figs. 2(b) and 2(c) show the overlapped silhouettes $\sigma_1 = \left(Y + v_1^1, G + v_1^2\right)$ and $\sigma_2 = \left(G + v_2^1, Y + v_2^2\right)$, respectively, for different motions (v_1^1, v_1^2, v_2^1 and v_2^2). Figs. 2(d) and (e) are the regions covered by the overlapped silhouettes, (i.e., H_{σ_1} and H_{σ_2}), with respect to Figs. 2(b) and (c), respectively.

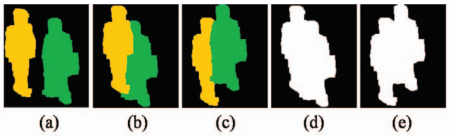

Fig. 2. An example of the overlapped silhouettes. (a) represents two silhouettes. The left yellow-silhouette is named Y and the right green-one is called G. (b) and (c) show two examples of the overlapped silhouettes σ_1 and σ_2 with respect to different motion speeds and orders, respectively. (d) and (e) show H_{σ_1} and H_{σ_2}, respectively.

3.1. Hypothesis Generation by PF

The state in our work is represented as an ordered sequence defined above. Given the observation at frame t, O_t, we would like to infer the most likely overlapped silhouettes, σ_t^*, that generates O_t. Let $\underline{O_t} = (O_1, O_2, ..., O_t)$ denotes the history of observations from the first frame to the t-th frame. PF uses a set of samples, or particles, $\{\sigma_t^i\}_{i=1}^N$, to approximate the posterior distribution $p\left(\sigma_t | \underline{O_t}\right)$ defined as follows:

$$p\left(\sigma_t \mid \underline{O_t}\right) \approx \sum_{i=1}^{N} w_t^i \delta(\sigma_t - \sigma_t^i), \tag{1}$$

where δ is a Dirac-delta function. In the bootstrap filter, the state transition probability is used as the important distribution, so that the associated un-normalized weight $\tilde{w}_t^i$ satisfies $\tilde{w}_t^i \propto \tilde{w}_{t-1}^i p\left(O_t | \sigma_t^i\right)$, where $p\left(O_t | \sigma_t^i\right)$ is the likelihood function. More details of PF can be referred to [9].

Assume that $\sigma_{t-1}^i = \{s^1, s^2, ..., s^M\}$, to propagate the state of the particle, the transition probability $p\left(\sigma_t^i | \sigma_{t-1}^i\right)$ is defined by the random-walk motion as follows:

$$\sigma_t^i = \left(s^1 + v_i^1, s^2 + v_i^2, ..., s^M + v_i^M\right), \tag{2}$$

where

$$v_i^k = (v_x, v_y)_i^k \sim N(0, \Sigma), \Sigma = diag\left(Var, Var\right), \tag{3}$$

and Var is a constant variance.

3.2. Likelihood

The likelihood $p\left(O_t | \sigma_t^i\right)$ employed in this work is adaptive to the occlusion situations as introduced below.

3.2.1. Binary-Silhouette Likelihood

The binary-silhouette likelihood measures the shape similarity between the O_t and σ_t^i based on difference and intersection sets, as shown in Fig. 3. Let $H = H_{\sigma_t^i}$ be the region covered by the overlapped silhouettes. The binary-silhouette likelihood $p_b\left(O_t | \sigma_t^i\right)$ is defined as follows:

$$p_b\left(O_t | \sigma_t^i\right) \propto \frac{f_b\left(R1\right)}{w_t^\sigma . f_b\left(R2\right) + w_t^O . f_b\left(R3\right)}, \tag{4}$$
$$R1 = H \cap O_t, R2 = H \cap \overline{O_t}, R3 = \overline{H} \cap O_t,$$

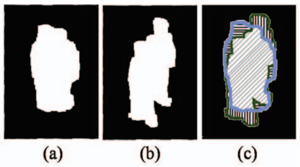

Fig. 3. The likelihood of binary silhouette. (a) presents the binary-silhouette of the O_t. (b) shows the binary-silhouette of the σ_t^i. (c) is the binary-silhouette likelihood. The area with slash line is R1, the area with vertical line is R2, and the area with horizontal line is R3.

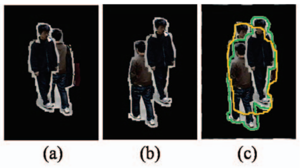

Fig. 4. The likelihood of color silhouette. (a) presents the color-silhouette of the O_t. (b) is the color-silhouette of the σ_t^i. (c) shows the color-silhouette likelihood. The yellow contour stands for the O_t, and the green contour represents the σ_t^i.

where $f_b\left(R\right)$ denotes the number of pixels within the region $R \in \{R1, R2, R3\}$. $\overline{H}$ and $\overline{O_t}$ are the complement sets of H and O_t, respectively. w_t^σ and w_t^O are the weights balanced between the hypothesis and the observation. In our implementation, $w_t^\sigma = w_t^O = 1$. Hence, (4) can be simplified as:

$$p_b\left(O_t|\sigma_t^i\right) \propto f_b\left(R1\right) f_b\left(R2 \cup R3\right)^{-1}. \qquad (5)$$

3.2.2. *Color-Silhouette Likelihood*

As shown in Fig. 4, the color-silhouette likelihood is measured based on both the shape and color-appearance similarities between O_t and σ_t^i. Let $f_c\left(R1\right)$ be the following measure that computes a normalized difference between the colors of O_t and σ_t^i in $R1$,

$$f_c\left(R1\right) \propto \frac{\sum_{x',y' \in R1} \|\sigma_t^i(x',y') - O_t(x',y')\|^2}{\sqrt{\sum_{x',y' \in R1} \|\sigma_t^i(x',y')\|^2 \sum_{x',y' \in R1} \|O_t(x',y')\|^2}}, \qquad (6)$$

where $\sigma_t^i(x',y')$ and $O_t(x',y')$ are the RGB colors at the pixel (x',y') of σ_t^i and O_t, respectively. Then, the color-silhouette likelihood $p_c\left(O_t|\sigma_t^i\right)$ is defined as

$$p_c\left(O_t|\sigma_t^i\right) \propto f_c\left(R1\right)^{-1} f_b\left(R2 \cup R3\right)^{-1}. \qquad (7)$$

3.2.3. *Switched-Silhouette Likelihood*

The binary-silhouette likelihood has less computation cost than that of the color-silhouette likelihood, but it cannot discriminate heavy occlusions, as shown in Fig. 1 column t_3 and t_4. To deal with heavy occlusions, the color-silhouette likelihood enumerates all the initial silhouette's permutation. Hence, a switched-silhouette likelihood is proposed to handle the slight and heavy occlusions accordingly. The switched-silhouette likelihood $p_s\left(O_t|\sigma_t^i\right)$ is defined as:

$$p_s\left(O_t|\sigma_t^i\right) = \begin{cases} p_c\left(O_t|\sigma_t^i\right) & \text{if } d \geq d_t \\ p_b\left(O_t|\sigma_t^i\right) & \text{otherwise} \end{cases} \qquad (8)$$

where d_t is the occlusion threshold and $d = \frac{S_{t1}}{O_t}$ measures of occlusion situation determined by the ratio of the initial silhouettes and the O_t. In our implementation, $d_t = 1.2$. Finally, σ_t^* is the tracked result of pedestrians' locations and is estimated as follows:

$$\sigma_t^* = \arg\max_{i=1}^{N} p_s\left(O_t|\sigma_t^i\right). \qquad (9)$$

4. EXPERIMENTAL RESULTS

To validate the proposed BATracker, we compare our method with Meanshift [10], Template Matching (Template) [11] and MILTracker [2] on three indoor video sequences[1], Seq1, Seq2 and Seq3. In our consideration, the background in all video sequences are removed in advanced. The number of particles applied to BATracker is set as $N = 30$ per silhouette. To evaluate the tracking performance, we use the mean position error (MPE) defined as follows:

$$MPE = \frac{\sum_{s \in S} |T(s) - G(s)|}{N_S}. \qquad (10)$$

where $|.|$ denotes $L1$-norm and N_S indicates the number of pedestrians in the pedestrian set S. $T(s)$ and $G(s)$ indicate the tracked 2D position and the ground truth position of the pedestrian s, where the ground truth is labelled manually.

The MPE results for each method are shown in Table. 1, and the numbers of frames in all sequences are shown as well. Meanshift relies on matching target's color histogram and Template searches for the most similar region, to avoid to mis-update the online information, both methods do not update the target's appearance model during the period of occlusions. However, the intrinsic variances and severe occlusions cause serious miss-tracking of the target. MILTracker drifts when severe occlusions or scale variances occur due to mis-updating the online information. By cooperating the silhouette combination and PF, the proposed BATracker successfully tracks the target even under severe occlusion situations. By switching binary and color-appearance silhouettes adaptively, BATracker locates the target efficiently. The snapshots of tracking results of Seq1, Seq2 and Seq3 are shown in Fig. 5, Fig. 6 and Fig. 7, respectively. In these figures, (a), (b), (c) and (d) are the outcomes of MILTracker, Meanshift, Template and BATracker, respectively. Because BATracker successfully locates the pedestrians' positions, the MPE performance of BATracker outperforms all of the compared tracking methods even under serious occlusions.

[1] Demo and original videos are available at: http://imp.iis.sinica.edu.tw/ivclab/research/batracker/index.html

Table 1. The compared MPE performances with different video sequences and different tracking methods.

	Seq1	Seq2	Seq3
# frames	239	56	220
MILTracker	18.27	56.82	34.22
Meanshift	23.46	24.92	27.51
Template	16.56	41.24	16.45
BATracker	3.49	4.39	4.24

5. CONCLUSION

In this paper, we proposed an approach to infer pedestrians' locations by considering binary-color silhouette similarity with PF. To achieve better tracking performance, the combination of silhouettes and the observation are considered to track pedestrians under different occlusion situations. For better tracking efficiency, the binary and color silhouettes are switched adaptively. The experimental results showed that BATracker outperforms the existing tracking methods we adopted for comparison. In the future, the scale and illumination variances will be considered in our approach.

6. ACKNOWLEDGEMENTS

This work was supported in parts by the National Science Council and Ministry of Economic Affairs, Taiwan, R.O.C., under Grant No. 98-2221-E-002-127-MY3 and No. 98-EC-17-A-02-S1-032, respectively.

7. REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM CS*, vol. 38, no. 4, pp. 13, 2006.

[2] B. Babenko, M. H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *CVPR*, 2009.

[3] Y. Wu, J. Cheng, J. Wang, and H. Lu, "Real-time visual tracking via incremental covariance tensor learning," in *ICCV*, 2009.

[4] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, "Robust tracking-by-detection using a detector confidence particle filter," in *ICCV*, 2009.

[5] W. Hu, X. Zhou, M. Hu, and S. Maybank, "Occlusion reasoning for tracking multiple people," *CSVT*, vol. 19, no. 1, pp. 114–121, 2009.

[6] L. Dong, V. Parameswaran, V. Ramesh, and I. Zoghlami, "Fast crowd segmentation using shape indexing," in *ICCV*, 2007.

[7] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *PAMI*, vol. 30, no. 7, pp. 1198–1211, 2008.

[8] Y. T. Chen, C. S. Chen, C. R. Huang, and Y. P. Hung, "Efficient hierarchical method for background subtraction," *PR*, vol. 40, no. 10, pp. 2706–2715, 2007.

[9] A. Doucet, N. D. Freitas, and E. N. Gordon, *Sequential Monte Carlo methods in practice*, 2001.

[10] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *PAMI*, vol. 25, no. 5, pp. 564–575, 2003.

[11] H. Schweitzer, J. W. Bell, and F. Wu, "Very fast template matching," in *ECCV*, 2002.

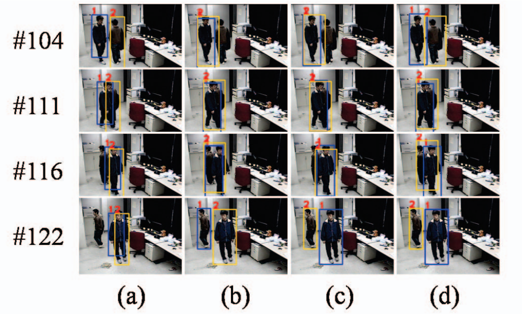

Fig. 5. The compared tracking result in Seq1. Each row illustrates the snapshots #104, #111, #116 and #122.

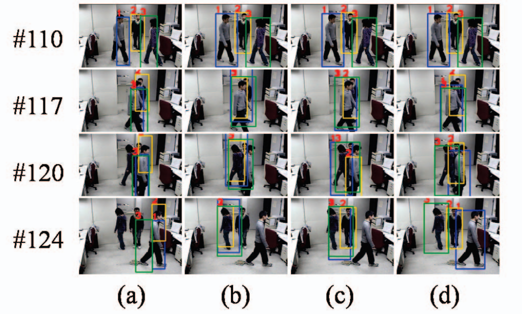

Fig. 6. The compared tracking result in Seq2. Each row shows the snapshots #110, #117, #120 and #124.

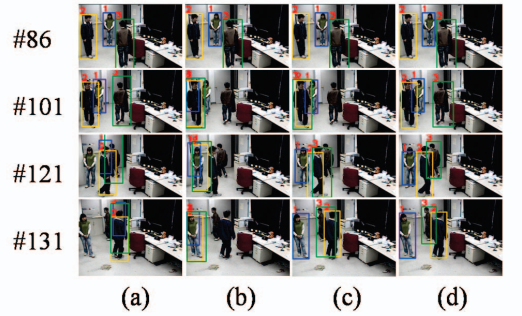

Fig. 7. The compared tracking result in Seq3. Each row contains the snapshots #86, #101, #121 and #131.