An Adaptive Approach for Overlapping People Tracking Based on Foreground Silhouettes

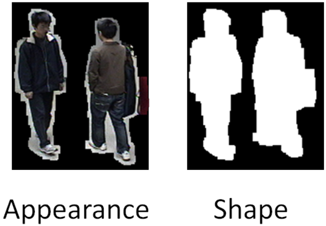
Hsin-Ho Yeh, Jiun-Yu Chen, Chun-Rong Huang and Chu-Song Chen
Institute of Information Science,
Acamedic Sinica,
Taiwan
{hhyeh, jychen, nckuos, song} @iis.sinica.edu.tw

2010/09/29

Outline

Introduction

- Pedestrian tracking is an important task in vision-based surveillance application.
- One of the difficulty in tracking problem is called feature variation
  - Appearance and shape are widely used in computer-vision applications.
  - Appearance is easily affected by illumination changes.
  - Shape suffers from the intrinsic variance, such as pose changed.



Introduction

- Another difficulty in tracking problem is called occlusion
  - Motion model is employed to help the tracker to locate the pedestrians.

Observed in frame t | Observed in frame t+1 | Motion prior

Related Works

- Constant motion model[1] cannot handle crowded environment.
- M.D. Breitenstein *et. al.*[2] incorporates pedestrian detector, online classifier and particle filter to infer the pedestrian locations, but it requires prior knowledge of human model.
- W. Hu *et. al.*[3] uses the ellipse shape, appearance model and particle filter to deduce the occlusion relationships between multiple people, but the ellipse shape model cannot handle human shape well due to pose variation.

---

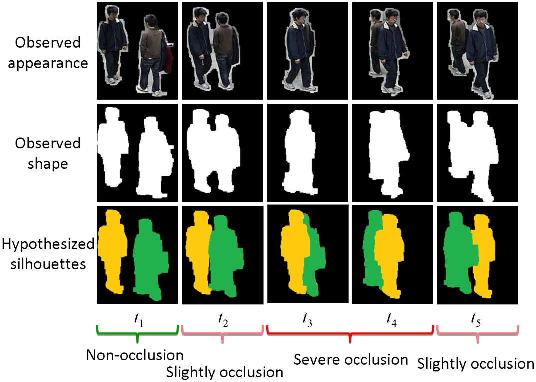
[1] B. Babenko, M. H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," CVPR, 2009

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

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

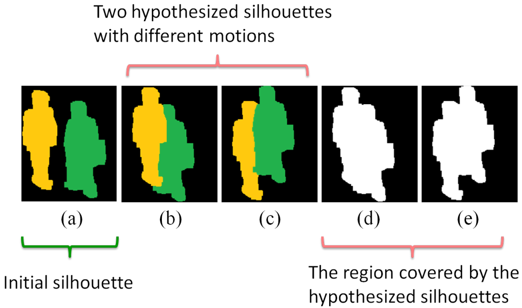
Motivation

- From above limitations, we propose a silhouette-based tracking method, called **B**inary/**A**ppearance **Tracker**, to handle the occlusion problem in pedestrian tracking.



Motivation - Intuition on Occlusion Handle

- To handle different occlusion situations,
  - Particle filter without motion prior is used to simulate the motion of hypothesis silhouettes.
  - The occlusion situation is inferred by the likelihood between the observed and hypothesis silhouettes.



BATracker - Hypothesis Generation

Particle filter(PF) uses a set of samples, $\{\sigma_t^i\}_{i=1}^N$, to approximate the posterior distribution 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, $\underline{O_t} = (O_1, O_2, ..., O_t)$ denotes the history of observations from first to t-th frame. 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 $k < l$.

BATracker - Hypothesis Generation

In bootstrap filter[4], the associated un-normalized weight $\tilde{w}_t^i$ is satisfied $\tilde{w}_t^i \propto \tilde{w}_{t-1}^i p\left(O_t|\sigma_t^i\right)$, where $p\left(O_t|\sigma_t^i\right)$ is the likelihood function. 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 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 $s^k + v_i^k$ stands for shifting the silhouette s^k by a 2D vector v_i^k and

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

Var is a constant variance.

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

BATracker - Binary-Silhouette Likelihood $p_b\left(O_t|\sigma_t^i\right)$

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

$$R1 = H \cap O_t, R2 = H \cap \overline{O_t}, R3 = \overline{H} \cap O_t,$$

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.

Observed shape | Hypothesized shape | Shape likelihood

BATracker - Color-Silhouette Likelihood $p_c\left(O_t|\sigma_t^i\right)$

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

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

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

Observed appearance | Hypothesized appearance | Appearance likelihood

BATracker - Switched-Silhouette Likelihood $p_s\left(O_t|\sigma_t^i\right)$

	Computational cost	Discriminateness
$p_b\left(O_t\|\sigma_t^i\right)$	O	X
$p_c\left(O_t\|\sigma_t^i\right)$	X	O

To incorporate both of discriminateness and low-computational cost, the switched-silhouette likelihood is defined as

$$p_s\left(O_t|\sigma_t^i\right) = \begin{cases} p_c\left(O_t|\sigma_t^i\right) & \text{if } d \geq d_t \\ p_b\left(O_t|\sigma_t^i\right) & \text{otherwise} \end{cases} \tag{4}$$

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, $d_t = 1.2$.

Experimental Results

- Dataset: Three indoor video sequences[5].
- Evaluation metric (Mean-Position Error):

$$MPE = \frac{\sum_{s \in S} |T(s) - G(s)|}{N_S}. \quad (5)$$

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.

[5] http://imp.iis.sinica.edu.tw/ivclab/research/batracker/index.html

Experimental Results

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

	Seq1	Seq2	Seq3
# frames	239	56	220
MILTracker[6]	18.27	56.82	34.22
Meanshift[7]	23.46	24.92	27.51
Template[8]	16.56	41.24	16.45
BATracker	3.49	4.39	4.24

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

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

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

Conclusions

To conclude our works,

- In this paper, we proposed an approach to handle different occlusion situations by considering silhouettes similarity with PF.
- For tracking effectiveness, the combination of silhouette and the observed silhouette are considered to track pedestrians' locations.
- For tracking efficiency, the binary/color silhouettes are switched adaptively to track pedestrians' positions under different occlusion situations.
- In the experimental results, BATracker can outperform the existing tracking methods that we adopted for comparison.

Future Work

In the future,

- The binary and color likelihood can be cooperated into learning-based approach to reduce scale and illumination variances.

THANK YOU

Thanks for your attention.

Reference

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

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

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.

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

A. Doucet, N. D. Freitas, and E. N. Gordon,
"Sequential Monte Carlo methods in practice,"
2001.

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

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