8-20

MVA2007 IAPR Conference on Machine Vision Applications, May 16-18, 2007, Tokyo, JAPAN

Particle Filter Based Tracking for Crossing of Targets with Similar Pattern

Yuji IWAHORI
Faculty of Engineering
Chubu University
Kasugai 487-8501, Japan
iwahori@cs.chubu.ac.jp

Shinya OKADA
Faculty of Engineering
Chubu University
Kasugai 487-8501, Japan
okada@cvl.cs.chubu.ac.jp

Haruki KAWANAKA
Faculty of Info. Sci. and Tech.
Aichi Pref. University
Nagakute 480-1198, Japan
kawanaka@ist.aichi-pu.ac.jp

Shinji FUKUI
Faculty of Education
Aichi University of Education
Kariya 448-8542, Japan
sfukui@auecc.aichi-edu.ac.jp

Robert J. WOODHAM
Dept. of Computer Science
University of British Columbia
Vancouver, B.C. Canada V6T 1Z4
woodham@cs.ubc.ca

Abstract

Tracking is an important topic in computer vision and object recognition. Recently, a probabilistic approach using particle filters has been applied to track moving objects. This kind of approach often uses a color histogram to estimate a likelihood function for probabilistic tracking. When two similar objects cross each other in view, the likelihood becomes high for both. This often causes tracking to fail.

This paper proposes a new method to address the object crossing problem. The method estimates the object region, splits the region into horizontal zones, and calculates similarity based on each split region and each horizontal zone of the target. The new method makes the tracking of similar targets more robust when those targets cross. Results are demonstrated on real video sequences.

1 Introduction

A particle filter is one technique for motion tracking that is robust in the presence of occlusion and noise. Particle filters, also called Sequential Monte Carlo methods (SMC) [1], are a sophisticated model estimation technique based on simulation. They implement Bayesian filtering and are used to estimate Bayesian models.

A particle filter is a maximum posteriori estimation method based on past and the present observations. It can achieve robust tracking even when the observation distribution is non-Gaussian. It approximates the discrete probability density where the random variables are represented by many particles. Particle filters now are widely used in motion tracking as well as in speech recognition and in many other applications.

A particle filter approach can be combined with other algorithms to make overall performance even stronger. The authors recently have proposed a particle filter approach to track moving objects using multiple cameras. This involves the passing of a moving object from the view of one camera to another over a wide region [2]. Also included is a method to track robustly under uniform illumination change [3].

One previous particle filter approach uses color histograms [4] except that the edge strength is used to calculate the likelihood function. Color histogram approaches can fail to track crossing or partially occluded objects when the colors are similar. Additional information is required. One idea is to combine the likelihood distributions obtained from from each camera in a multiple camera and hence multiple view configuration [5]. Another is to impose specific spatio-temporal knowledge, for example, a model of human walking [6].

This paper describes a new approach to improve object tracking with a single camera and a fixed viewpoint. The method estimates the image region corresponding to the object tracked, splits the particles into sub-regions and calculates similarity to detect the interference of one object with another. The new method makes the tracking of similar targets more robust when those targets cross in view. Results are demonstrated on real video sequences.

2 Particle Filter Based Tracking

2.1 Principle

Time sequential filtering is a method to estimate the most suitable value from the past and present observation values. Let the state of tracking a target at time t be $\mathbf{x}_t$, and let the corresponding image observation be $\mathbf{z}_t$. Let the total observation results at time t be $\mathbf{Z}_t = (\mathbf{z}_1, \ldots, \mathbf{z}_t)$. The probability density of $\mathbf{x}_t$ given $\mathbf{z}_t$ is discretely approximated by many particles with state and likelihood. Tracking robust to both noise and variation in the environment is performed.

Particle filtering approximates the posterior $p(\mathbf{x}_t|\mathbf{Z}_t)$ at time t with N particles which consist of the state $\mathbf{x}$ and an associated weight. The weight $\pi_t^{(i)}$ for state $\mathbf{x}_t^{(i)}$, the i-th hypothesis, at time t is evaluated by the likelihood function $p\left(\mathbf{z}_t|\mathbf{x}_t = \mathbf{x}_t^{(i)}\right)$. Tracking with hypotheses is realized by repeating the following process.

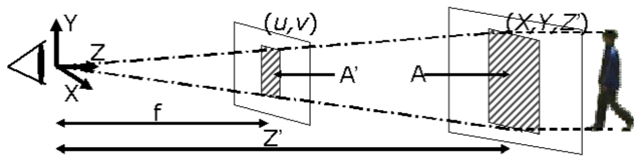

Figure 1. Weak Perspective Projection

1. Sample the hypotheses $\left\{\mathbf{s}'^{(1)}_{t-1}, \ldots, \mathbf{s}'^{(N)}_{t-1}\right\}$ using the weight $\pi^{(i)}_{t-1}$ and the state $\mathbf{x}^{(i)}_{t-1}$ of particles $\left\{\left(\mathbf{x}^{(i)}_{t-1}, \pi^{(i)}_{t-1}\right), i = 1, \ldots, N\right\}$ to approximate the posterior distribution $p\left(\mathbf{x}_{t-1}|\mathbf{Z}_{t-1}\right)$ at time $t-1$.

2. Generate N hypotheses $\mathbf{x}^{(i)}_t$ at time t from the sampled hypotheses $\mathbf{s}'^{(i)}_{t-1}$.

3. Estimate likelihood functions from the $\mathbf{x}^{(i)}_t$ and the weights $\pi^{(i)}_t$. Here, the weights are normalized so that $\sum_{i=1}^{N} \pi^{(i)}_t = 1$.

   Particles $\left\{\left(\mathbf{x}^{(i)}_t, \pi^{(i)}_t\right), i = 1, \ldots, N\right\}$ are obtained as a discrete approximation of the posterior distribution $p\left(\mathbf{x}_t|\mathbf{Z}_t\right)$ at time t. The mean value of the hypotheses is used as the estimated state for the tracked target at time t.

2.2 Estimation of Target Size

Estimating the target region to track improves tracking precision. As shown in Figure 1, weak perspective projection assumes that camera centric coordinates (X, Y, Z) project to image coordinates (u, v). A pinhole camera models gives

$$u = f\frac{X}{Z'} \qquad v = f\frac{Y}{Z'} \tag{1}$$

Z' is the distance, assumed fixed, of the object from the lens along the Z-axis. Let A be the area of the object projected onto the plane parallel to the image plane. Then the corresponding area A' on the image plane is estimated by

$$A' = f^2 \frac{A}{Z'^2} \tag{2}$$

If A, Z' and f are known *a priori*, then A' is an estimate of target size in the image. Similarly, the aspect ratio of A estimates target width and height. Camera calibration determines target size and aspect ratio. Coordinates (u, v) are used as state variables.

3 Crossing of Targets with Similar Pattern

There are two cases to consider at a crossing. The first case is when multiple particle groups track a single target. The second case is when particle groups switch targets.

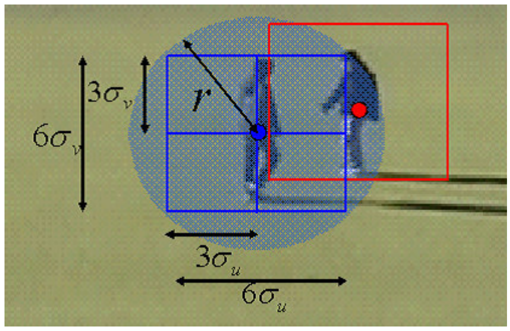

Figure 2. Circular Region for Detection of Interference

3.1 Interference of Particle Groups

In this section, a method to judge when two particle groups come close to each other is described. Particles are generated by the Box-Muller method with Gaussian distribution of random variables. Image coordinates (u, v) are the random state variables. Particles are distributed according to a 2D Gaussian distribution with constant values of the standard deviations σ_u and σ_v. For a Gaussian, about 99.74% of the particles are included within a window of $\pm 3\sigma$. We assume $\pm 3\sigma_u$ as the maximum width in the u-direction and $\pm 3\sigma_v$ as the maximum width in the v-direction.

Let (u_t, v_t) be the image coordinates of the tracked object estimated as the weighted mean value of all the particles. Consider a circle with radius r centered at (u_t, v_t) where

$$r = \sqrt{(3\sigma_u)^2 + (3\sigma_v)^2} \tag{3}$$

This circle with radius r is used to determine interference among particle groups. As shown in Figure 2, when the particle group of a second tracked object overlaps the circle associated with the particle group of a first, the two objects are judged to have become close to each other.

3.2 Horizontal Split Histogram

As shown in Figure 3, the rectangular tracking region is split into m regions horizontally. A hue (H) and saturation (S) histogram is generated for each region $H_i (i = 1, 2, 3..., m)$. The HS histogram of the first frame, also split into m regions horizontally, is used as a reference histogram. The similarity S_i for each region is calculated by Swain's histogram intersection [10] as

$$S_i = \sum_{u=1}^{T} \min(H_u, H_u^{ref}) \tag{4}$$

where $1, \ldots, T$ is the range of H and S values in the HS histogram. The overall similarity S is obtained as the mean of the similarities S_i for each region.

$$S = \sum_{i=1}^{m} \frac{S_i}{m} \tag{5}$$

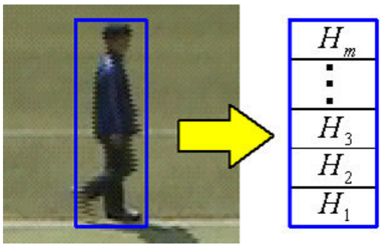

Figure 3. Horizontal Split Histogram

3.3 Weighting of Particles Using Distance

When two particle groups A and B come close to each other, the likelihood can become high for both tracked objects. As a result, one of them can become tracked by both particle groups. Figure 4 provides an example. The criterion for overlap is as discussed before. To improve tracking, the likelihood of a particle hypothesis $\mathbf{s}_t^{(i)}$ in group A is reduced when group B overlaps group A. Individual particle hypotheses belonging to group A are assigned smaller weights when they come close to group B and larger weights when they are more distant from group B. As without overlap, the weight assigned to individual particle hypotheses in group A is inversely related to the distance from group A.

This is implemented as follows. Let $d_1^{(i,ja)}$ be the distance between a particle i belonging to group A and another particle ja also belonging to group A. The mean distance $\overline{d_1^{(i)}}$ is calculated inside group A according to Eq.(6). Similarly, let $d_2^{(i,jb)}$ be the distance between a particle i belonging to group A and another particle jb belonging to group B. The mean distance $\overline{d_2^{(i)}}$ to group B is calculated according to Eq.(6). That is, for all particles in group A, $\overline{d_1^{(i)}}$ and $\overline{d_2^{(i)}}$ $(i = 1, \ldots, N)$ are given as

$$\overline{d_1^{(i)}} = \sum_{ja=1}^{N} \frac{d_1^{(i,ja)}}{N}, \quad \overline{d_2^{(i)}} = \sum_{jb=1}^{N} \frac{d_2^{(i,jb)}}{N} \tag{6}$$

Consider the likelihood function given in Eq.(8). The similarity S of the color histogram has a value between 0 and 1. We normalize the distance values, as well. Let $\overline{d_1}_{\max}$ and $\overline{d_2}_{\max}$ be the maximum values, respectively, of $\overline{d_1}$ and $\overline{d_2}$, the mean values of the distance to each group. Then, the normalized distances $D_A^{(i)}$ and $D_B^{(i)}$ are obtained as

$$D_A^{(i)} = \frac{\overline{d_1^{(i)}}}{\overline{d_1}_{\max}}, \quad D_B^{(i)} = \frac{\overline{d_2^{(i)}}}{\overline{d_2}_{\max}} \tag{7}$$

Recall that S is the similarity obtained from the color histogram. A standard likelihood function to track a single moving object is

$$L(\mathbf{z}_t|\mathbf{x}_t) = \exp(kS^2) \tag{8}$$

where k is a constant. The modified likelihood function used here is

$$L(\mathbf{z}_t|\mathbf{x}_t) = \exp(k_1 S^2)\exp(k_2 D_B^{(i)}(1 - D_A^{(i)})) \tag{9}$$

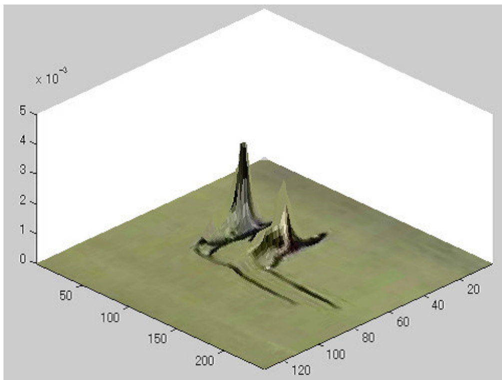

Figure 4. Likelihood Distribution in Case of Tracking Left Object

where k_1 and k_2 are constants.

4 Experiments

The implementation uses C++ on MS Windows XP and DirectShow. The video camera is a SONY DCR-HC40. Frames in the video sequences are 720 $\times$ 480 pixels, 24 bit color. The PC used was an Athlon64$\times$2 Dual Core Processor 4200+ with 2 GB Main Memory.

Comparison is between a previous approach [4], based on Eq.(8), and the proposed approach, based on Eq.(9). The number of particles for each target is 50. Constants k, k_1 and k_2 are set to 20, 20 and 5, respectively. Success ratios and failure ratios for object crossings are calculated. “Success” is when the particles track each target correctly during a crossing.

There are two failure cases to consider. The first case is when the tracked target changes with the crossing (Failure A). The second case is when multiple particles track a single target after the crossing (Failure B). Four kinds of video sequences were evaluated. Examples from scene 1 and scene 3 are shown in Figure 5 and 6, respectively. Significant improvement is observed. Evaluation results for four scenes are shown in Table 1 for [4] and in Table 2 for the proposed approach.

Table 1. Evaluations for Previous Approach

Scene	Succ.Ratio [%]	Fail.A [%]	Fail.B [%]
1	65.22	0.00	34.78
2	80.00	0.00	20.00
3	7.50	0.00	92.50
4	77.19	0.00	22.81

Table 2. Evaluations for Proposed Approach

Scene	Succ.Ratio [%]	Fail.A [%]	Fail.B [%]
1	100.00	0.00	0.00
2	100.00	0.00	0.00
3	88.57	0.00	11.43
4	58.95	40.00	1.05

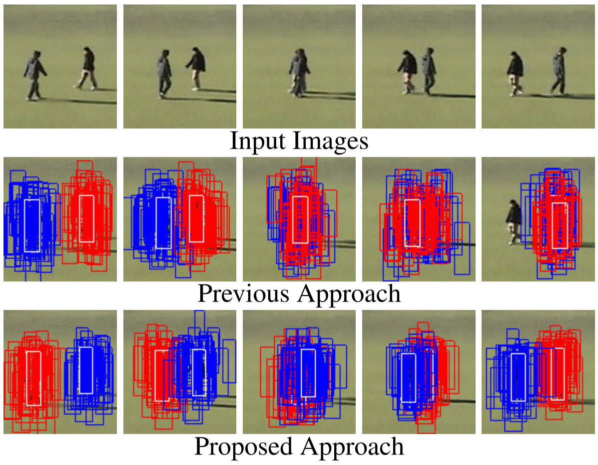

Figure 5. Examples of Scene 1

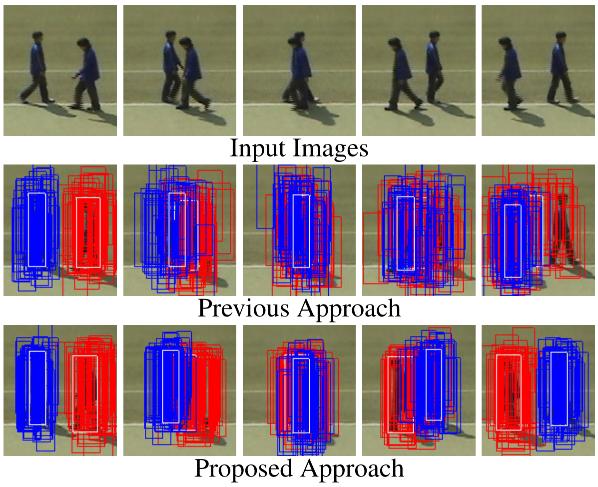

Figure 6. Examples of Scene 3

Table 1 and Table 2 suggest that the ratio of Failure B becomes lower for all scenes. This demonstrates that separating the particle groups based on the distance of particles can make tracking crossings of similar targets more robust. The success ratio also increases for scenes 1 to 3. For scene 4, the ratio of Failure B decreases but the ratio of Failure A increases resulting in an overall decrease in the success ratio. The increase in Failure A resulted from additional target switches even when the two particle groups did not come very close each other. When this happens, it remains necessary to correct tracking by other means. In our implementation, the previous approach took 9.7 [ms/f] and the proposed approach took 10.0 [ms/f]. In each case, this exceeds the video rate of 30 [fps] and real-time performance is achieved.

5 Conclusion

This paper presented a new approach to object tracking that is robust even when visually similar targets cross in view. The method uses a similarity measure calculated from the horizontal split histogram and a distance measure to the particle groups. Tracking when objects cross is better able to keep particle groups separate, since overlap is made explicit and distance within and between particle groups is incorporated into the likelihood function. Experiments confirm that cases where two distinct particle groups converge to a single target are reduced. Cases involving more than two targets crossing simultaneously remain as a topic for further study.

Acknowledgment

Iwahori's research is supported by the Grant-in-Aid for Scientific Research No.16500108, Japan Society for the Promotion of Science and Chubu Univ. Grant. Woodham's research is supported by the Canadian Natural Sciences and Engineering Research Council (NSERC).

References

[1] M. Isard and A. Blake: "Condensation – conditional density propagation for visual tracking", Intl. J. Computer Vision, Vol.29, pp.5–28, 1998.

[2] T. Takai, Y. Iwahori, H. Kawanaka, H. Itoh: "Tracking of Moving Objects Using Multiple Fixed Camera", Proceedings of SSII05 (11th Symposium on Sensing via Image Information), pp.189–192, 2005 (in Japanese).

[3] Y. Iwahori, H. Kawanaka, T. Takai, Y. Adachi, H. Itoh: "Particle Filter Based Tracking of Moving Object from Image Sequence", Springer LNCS, Volume 4252/2006, pp.401–408, KES2006.

[4] P. Perez, C. Hue, J. Vermaak and M. Gangnet: "Color-Based Probabilistic Tracking", Proc. ECCV 2002, Part I, pp.661–675, 2002.

[5] S. Khan and M. Shah: "A Multiview approach to tracking people in crowded scenes using a planar homography constraint", Proc. 9th ECCV, Vol.4, pp.133–146, 2006.

[6] Tao Zhao, Ram Nevatia, Fengjun Lv: "Segmentation and Tracking of Multiple Humans in Complex Situations", Proc. CVPR 2001, Vol.2, pp.194–201, 2001.

[7] Y. Satoh, T. Okatani and K. Deguchi: "A Color-based Tracking by Kalman Particle Filter",Proc. ICPR 2004, Vol. 3, pp. 502–505, 2004.

[8] J. Sherrah and S. Gong: "Fusion of perceptual cues for robust tracking of head pose and position", Pattern Recognition, Vol. 34, No.8, 2001.

[9] A. Doucet, S. Godsill and C. Andrieu: "On sequential Monte Carlo sampling methods for Bayesian filtering", Statistics and Computing, Vol.10, pp.197–208, 2000.

[10] M. J. Swain and D. H. Ballard: "Color indexing", Intl. J. Comp. Vision, Vol.7, pp.11–32, 1991.