International Journal of Signal Processing, Image Processing and Pattern
Vol. 2, No.1, March, 2009

Video Based Moving Object Tracking by Particle Filter

Md. Zahidul Islam, Chi-Min Oh and Chil-Woo Lee
Chonnam National University, Gwangju, South Korea
zahid@image.chonnam.ac.kr

Abstract

Usually, the video based object tracking deal with non-stationary image stream that changes over time. Robust and Real time moving object tracking is a problematic issue in computer vision research area. Most of the existing algorithms are able to track only in predefined and well controlled environment. Some cases, they don't consider non-linearity problem. In our paper, we develop such a system which considers color information, distance transform (DT) based shape information and also nonlinearity. Particle filtering has been proven very successful for non-gaussian and non-linear estimation problems. We examine the difficulties of video based tracking and step by step we analyze these issues. In our first approach, we develop the color based particle filter tracker that relies on the deterministic search of window, whose color content matches a reference histogram model. A simple HSV histogram-based color model is used to develop this observation system. Secondly, we describe a new approach for moving object tracking with particle filter by shape information. The shape similarity between a template and estimated regions in the video scene is measured by their normalized cross-correlation of distance transformed images. Our observation system of particle filter is based on shape from distance transformed edge features. Template is created instantly by selecting any object from the video scene by a rectangle. Finally, in this paper we illustrate how our system is improved by using both these two cues with non linearity.

1. Introduction

Video based moving object tracking is one of the exigent missions in computer vision area such as visual surveillance, human computer interactions etc. Video based tracking basically accord with non-stationary image, target object descriptions and the background which change over time. The most available algorithms are able to perform tracking simply in predefined and well controlled environment. In some cases, they don't consider the non-linearity problem. Actually, without this dynamic performance the system can not be applied in real time. Our goal is to make a real time tracking system and in this case we carefully consider the color, shape and non-linearity. Tracking objects is performed in a sequence of video frames and its process mainly consists of two main stages: isolation of objects from background in each frames and association of objects in successive frames in to trace them. According to the most of existing system, they are able to track object in an image sequence, either viewed or not or based on some extra trained data, in dumpy period and in well controlled environment. These algorithms usually fail to perfectly observe the deformable object shape changing in video images and in the bulky lighting (illumination) variations. In our developed algorithm we all consider these kinds of problems.

To start object tracking, generally the trackers need to be initialized by an external module [1]. Object model for tracking in image processing is usually based on reference image of the object, or properties of the objects [5]. Once an object model is initiated the tracking algorithms will conduct based on high correlations of the object motion, shape, color or appearance from the model between consecutive video frames. But, unfortunately robust and efficient object tracking is still an open research issue and in this paper our approach also to present an efficient system to make a robust tracker. In our present first case, we use HSV histogram-based object model and particle filter to make a robust color based probabilistic tracker system. For the second case, we proposed DT template object model and new similarity measurement method that is a new DT template based method. In this method, to measure the similarity between the template and a video image at estimated object regions from particles, we use distance transform (DT) edge template matching. To track efficiently we need a good observation model and we demonstrate normalized cross-correlation based observation model doing well for object tracking. And finally, we achieve an immense result using these two features concurrently in a same video sequences.

The remaining part of this paper is organized is as follows: section 2 describes related and existing work in this field that motivated this work. Section 3 introduces basic particle filter for estimation problem and object tracking. And then section 4 describes color histogram distance based system. The proposed system models which unified distance transform, and normalized cross-correlation with particle filter, are discussed in section 5. In this section we also present some comparison result of the proposed system. The numerous experiments and performance evaluation are also presented here. Conclusive remarks are addressed at the end of the paper in section 6.

2. Related works and motivation

In this section we focus on various models and techniques for video based object tracking. We review here only the most relevant video based tracking, which is based on shape like contour, edge and color with particle filter. Traditionally, the tracking problem is formulated as a sequential recursive estimation having on an estimate of the probability distribution of the target in the previous frame; the problem is to estimate the target distribution in the new frame using all available prior knowledge and new information brought by the new frame. The state-space formalism, where the current tracked object properties are described in an unknown state vector updated by noisy measurements, is very well adapted to model tracking. Unfortunately the sequential estimation has an analytical solution under very restrictive hypothesis. The well known Kalman filter is such a solution and is optimal for the class of linear Gaussian estimation problems. The Particle filter, a numerical method that allows finding an approximate solution to the sequential estimation, has been successfully used in many target tracking problems and visual tracking problems. But Kalman filter has limitations for multidimensional tracking. Particle filter success, in comparison with Kalman filter, can be explained by its capability to cope with multi-modality of the measurement densities and non-linear observation models. In visual tracking, multi-modality of the measurement density is very frequent due to the presence of multifaceted scene elements which has a similar appearance to the target. The observation model, which relates the state vector to the measurements, is non-linear because image data endures feature extraction, and a highly non-linear operation. This is the intention to work with particle filter to develop our video based tracking in terms of handling non-linearity.

For hand tracking according to Israd et al. [6], they apply CONDENSATION algorithm to combine skin color and hand contour. But that is not a general system for moving object tracking. There is an extension of CONDENSATION algorithm in [11].

But this algorithm can not be applicable for any all-purpose system and skin color has some limitations. For object tracking, a color based particle filter is proposed in many works [4,5]. In their case, target model of the particle filter is defined by the color information of the tracked object. According to Lehuger et al. [4], an adaptive mixture color model is used for updating reference color model to make more robust visual tracking. Some researchers try to integrate, multiple features such as color, shape, motion, edge. But finally they apply only color and texture cue in particle filter based implementation. Moving edge features is used in [3]. So, from the various literature reviews we observe that, it is still a tricky task for more robust real time object tracking. And our work is motivated to do with particle filter based efficient two dimensional DT image matching algorithm. Also, we then combine this approach with color to make it more effective in any environment.

3. Particle filtering

Usually, for any video based tracking system, our goal is to track object through a sequence of video. Tracking objects in video involves the modeling of non-linear and non-gaussian systems. If we have to consider these two phenomena, then how can we get solution for dynamic moving object tracking? One solution can be employed by using a probabilistic framework which formulates tracking as interference in a Hidden Markov Model (HMM). This frame is like as shown in figure 1. In order to model accurately the underlying dynamics of a physical system, it is important to include elements of non-linearity and non-gaussianity in many application areas. Particle Filters can be used to achieve this. They are sequential Monte Carlo methods based on point mass representations of probability densities, which are applied to any state model [2]. Particle Filter is a hypothesis tracker, which approximates the filtered posterior distribution by a set of weighted particles. It weights particles based on a likelihood score and then propagates these particles according to a motion model.

Weight of each particle should be changed depending on observation for current frame. The basic Particle Filter algorithm consists of 2 steps: Sequential importance sampling (SIS) and Selection step. In SIS step it uses Sequential Monte Carlo Simulation. For each particle at time t, transition priors are sampled. For each particle we then evaluate and normalize the importance weights. In selection steps (Resampling), we multiply or discard particles with respect to high or low importance weights to obtain a predefined number of particles. This selection step is what allows us to track moving objects efficiently.

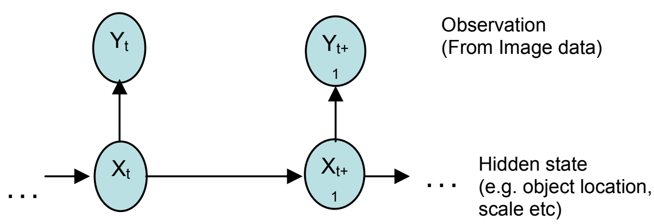

Figure 1. Probabilistic frame work for tracking system

Particle filter consists of essentailly two steps: prediction and update. Given all available observations $y_{1:t-1} = \{y_1, \ldots, y_{t-1}\}$ up to time $t-1$, the prediction stage uses the probabilistic system transition model $p(x_t \mid x_{t-1})$ to predict the posterior at time t as

$$p(x_t \mid y_{t-1}) = \int p(x_t \mid x_{t-1}) p(x_{t-1} \mid y_{1:t-1}) dx_{t-1} \quad (1)$$

At time t, the observation y_t is available, the state can be updated using Bay's rule

$$p(x_t \mid y_{1:t}) = \frac{p(y_t \mid x_t) p(x_t \mid y_{1:t-1})}{p(y_t \mid y_{1:t-1})} \quad (2)$$

where $p(y_t \mid x_t)$ is described by the observation equation.

In the particle filter, the posterior $p(x_t \mid y_{1:t})$ is approximated by a finite set of N samples $\{x_t^i\}_{i=1,\cdots,N}$ with importance weights w_t^i. The candidate samples $\tilde{x}_t^i$ are drawn from an importance distribution $q(x_t \mid x_{1:t-1}, y_{1:t})$ and the weight of the samples are –

$$w_t^i = w_{t-1}^i \frac{p(y_t \mid \tilde{x}_t^i) p(\tilde{x}_t^i) \mid x_{t-1}^i}{q(\tilde{x}_t \mid x_{1:t-1}, y_{1:t})} \quad (3)$$

The samples are resampled to generate an unweighted particle set according to their importance weights to avoid degeneracy. In the case of the bootstrap filter [13], $q(x_t \mid x_{1:t-1}, y_{1:t}) = p(x_t \mid x_{t-1})$ and the weights become the observation likelihood $p(y_t \mid x_t)$.

3.1. Mathematical Model

For implementation of particle filter we need the following mathematical model:

1. Transition model / state motion model $P(x_t \mid x_{t-1})$: this specifies how objects move between frames.
2. Observation model $p(y_t \mid x_t)$: this specifies the likelihood of an object being in a specific state (i.e. at the specific location).
3. Initial state $Est(1)$ / prior distribution model $p(x_0)$: describes initial distribution of object states.

4. Color based system

4.1. System flow diagram

We want to apply a particle filter in a color model based framework. This system depends on the deterministic search of a window, whose color content matches a reference histogram color model. We use principle of color histogram distance. The overall working flow diagram is shown in figure 2.

We have modeled the states, as it location in each frame of the video. The state space is represented in the spatial domain as X = (x,y). The state space has been initialized for the first frame manually.

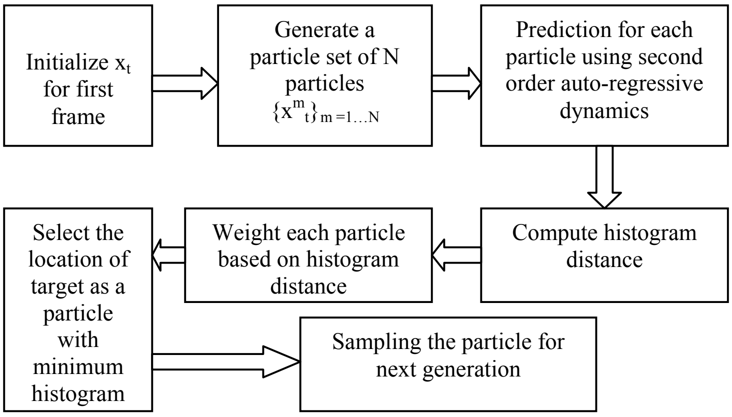

Figure 2. Color based particle filter implementation

4.3. Observation model and system dynamics

The observation y_t is proportional to the histogram distance between the color window of the predicted location in the frame and the reference color window. A second order auto-regressive dynamics is chosen on the parameters used to represent our state space that is, (x,y). The dynamics is given by $X_{t+1}=Ax_t+Bx_{t-1}$, where the matrices A and B could be learned from a set of sequences where correct tracks have been obtained.

4.3. Experiments result

In this section we will demonstrate some experimental results on several real-world video sequences captured by pan/tilt/zoom video camera in indoor, outdoor environment. Figure 3 and 4 show the effectiveness of our system in indoor and outdoor situations.

5. DT template shape based system

5.1. Preprocessing

The job of robust and real time tracking requires a robust observation model. The task of robust tracking demands a robust observation model. To make template based observation model, in this paper, first we select and extract object shape information from a video scene by edge detection (canny edge detector) and then employ robust shape matching based on distance transformation [8] and compares between first initialized object and tracked object by normalized cross-correlation [9].

Matching involves correlating the templates with the distance transformed scene and determining the locations where the mismatch is below a certain user defined threshold. Template is updated in adaptive fashion in tracking sequence. The preprocessing functional block diagram is shown in figure 5.

Figure 3. Object tracking sequence in indoor

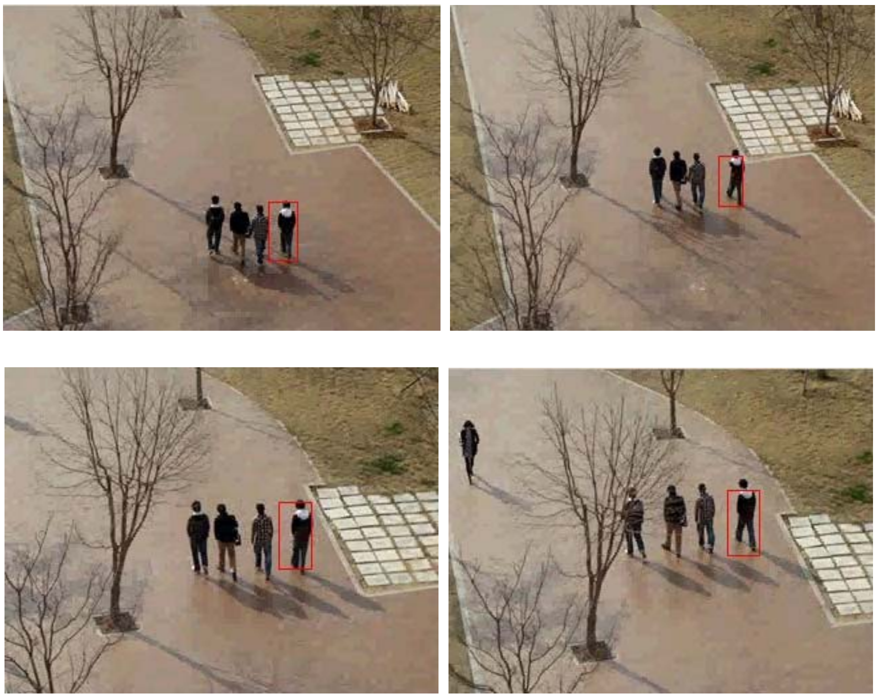

Figure 4. In outdoor object tracking sequence

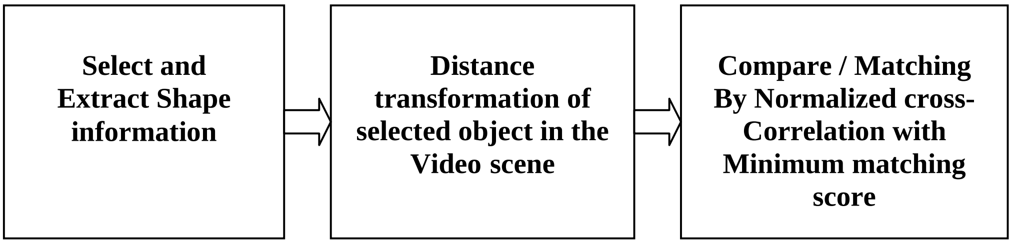

Figure 5. Preprocessing steps

5.2. Distance transformation (DT)

The typical matching with DT [10] is done between two binary images; a segmented template and a segmented image which we call feature template and feature image. To formalize the idea of DT matching similar, the shape of an object is represented by a set of point. The image map is represented as a set of feature points. In our present work we always update our reference template in every frame. So, it is more robust for matching in each changing of tracked object. The typical original image and DT image map is shown in figure 6. In the original image if we select the hand as template for matching to find it, followed by DT we can match it very smoothly and robustly by means of normalized cross-correlation which is discussed in next section.

5.2. Matching by normalized cross-correlation

Traditional correlation based matching methods are limited to the short baseline case [9]. According to Zhao et al. [9], normalized cross-correlation (NCC) is proposed for matching two images with large camera motion.

(a) (b)

Figure 6. (a) Original image (b) Corresponding DT image map

Their method is based on the rotation and scale invariant normalized cross-correlation. In our present case, we use normalized cross-correlation for image processing applications in which the brightness of the image and template can vary due to lighting and exposure conditions, the image can be first normalized. This is done at every step by subtracting the mean and dividing by the standard deviation. That is, the cross-correlation of template t(x,y) with a subimage f(x,y) is given by the following equation

$$N_{f,t} = \sum_{x,y} \frac{(f(x,y) - \bar{f})(t(x,y) - \bar{t})}{\sigma_f \sigma_t} \tag{4}$$

5.2. Particle filter based implementation

In our present proposed system, we integrate the template based shape matching by normalized cross-correlation with particle filter for robust object tracking. In our implementation we flow some steps which is shown in figure 7.

Steps:

- **Adaptive template based probabilistic tracking**
  - Deterministic search, shape information matches a reference templa te
- **State Space**
  - The state space is represented in the spatial domain as X = (x,y)
  - Initialized state space for the first frame manually by selecting the o bject of interest in the video scene by rectangle.
- **System Dynamics**
  - A second-order auto regressive dynamics is chosen on the paramet ers used to represent our state space.
- **Observation y_t**
  - The observation y_t is proportional to the normalized cross-correlatio n matching score between tracked window of the predicted location i n the frame and the reference template.
  - y_t α NCC Score (q, q_x), where, q is the reference template and q_x is t he predicted location window

Figure 7. Particle filter based implementation

The algorithm flowchart of the particle filter iteration using our proposed DT template based method is shown in figure 8.

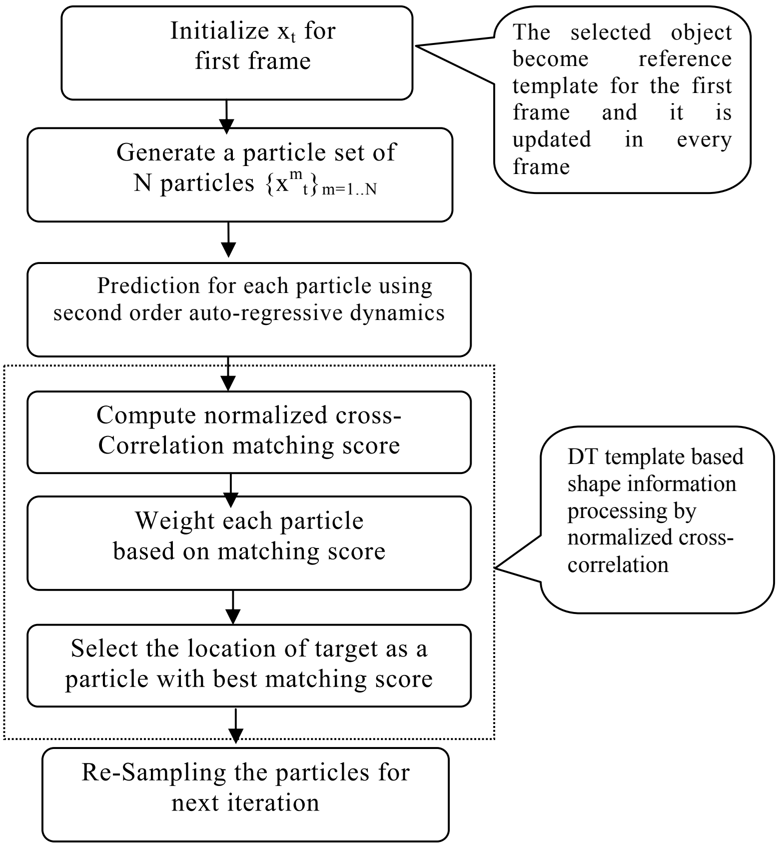

Figure 8. Algorithm flow chart for shape based system

5.5. Experimental results

The cross correlation based template matching is motivated by the distance measure (squared Euclidean distance). There are several disadvantages using this approach, for example, if the image energy varies with position, matching can fail. Moreover, it is not invariant to changes in image amplitude such as those caused by changing lighting conditions across the image sequences. The correlation coefficient overcomes these difficulties by normalizing the image and feature vectors to unit lengths, yielding a cosine-like correlation coefficient. So we are intended in our present work for DT image map based matching with normalized cross-correlation to develop observation model to make more robust particle filter based tracker. We try to take both advantages from DT image map and normalized cross correlation regarding to develop our particle filter based system. In this section we will present some experimental results on several real-world video sequences captured by pan/tilt/zoom video camera. Our all resulting image sequences are presented here from left to right, the top left corner number shows the frame number.

It is same as in color based system in various environments in our previous experiments. For all testing sequences, we use the same algorithm configuration, e.g. state-space, system dynamics, observation model and template based probabilistic tracking. The first sequences as shown in figure 9, shows tracking a single moving person from top view.

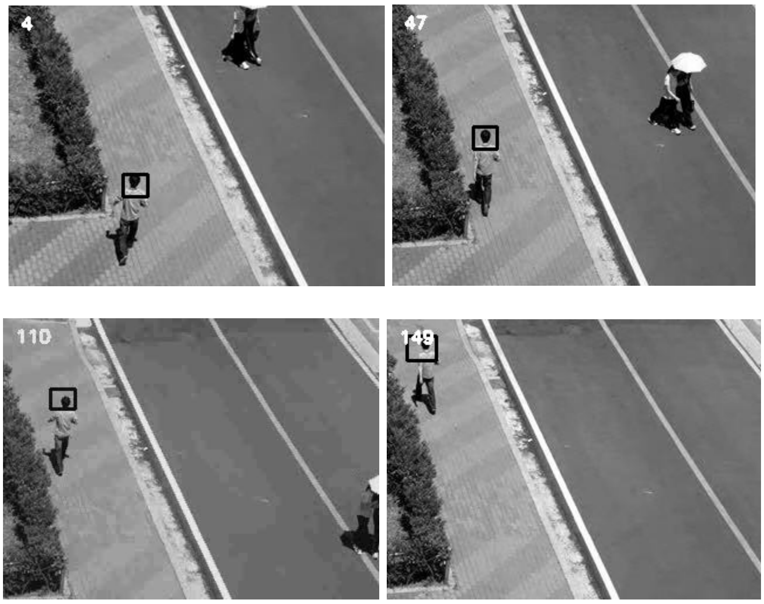

Figure 9. First sequence from top view

Figure 10. Second resulting sequence in a play ground

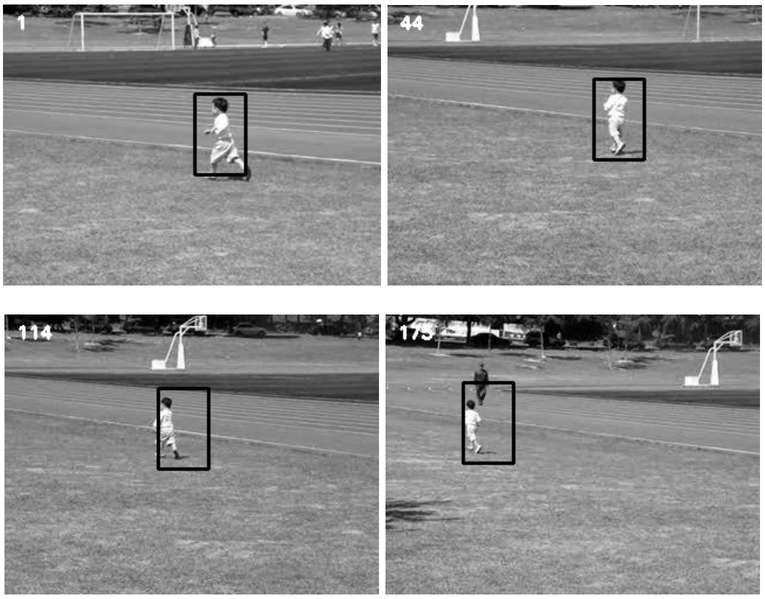

Figure 11. Third resulting sequence in a play ground

The second and third resulting sequences as shown in figure 10 and 11 are also showing the effectiveness of our proposed system. These sequences are taken from a playground with cluttered background for tracking single moving person. In the all figures the top left most white text is the frame number.

For evaluate our proposed tracker with different features conditions, we calculate the RMS error of feature tracking point. We compare our result with the manually-labeled "ground truth" locations of the features. The figure 12 shows the RMS error between ground truth reference points with target point. The blue curve is the result of considering only edge features.

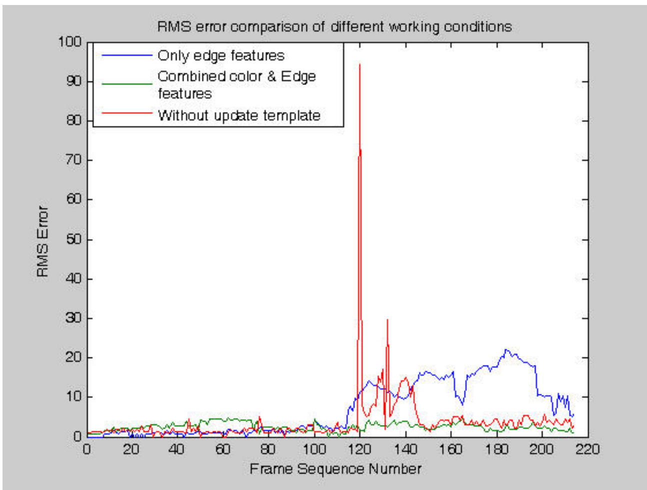

Figure 12. The RMS error at tracking feature points for each frame

The green curve of this figure is combined effect of edge and color features. This shows better result than only edge features. In our proposed system, we always, update the DT template in every frame. The red curve shows the performance of the system without update template. In this case, we see, in some frame, its highly drifted from the target locations and unstable. We also observe that, our system become unstable due to start camera zoom at frame number 115. But the combined effect of color and DT edge features result shows the outstanding performance (green curve). Figure 13 shows the samples distribution considering the best matching score. This probability density function (pdf) represents the best match to the target object.

In our all experiments we keep constant the particle number at only 100. And each video consists of 320 x 320 pixel images with 30 fps. This tracking algorithm is implemented in C++ and OpenCV library on Windows XP platform with standard Pentium 4 (with 1.5 GB RAM) machine. These all results show the algorithm performance under different scenarios.

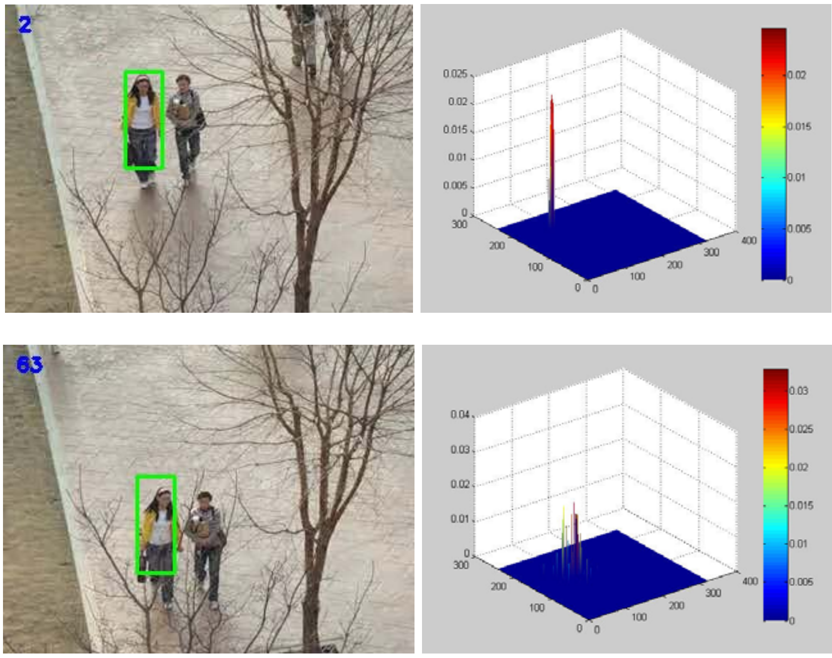

Figure 13. Probability density function corresponding to target object

8. Conclusion

In this paper, two techniques have been presented for tracking a single moving object in a video sequence using color and shape information. We also, carefully examine the combined effect of these to features which performance is better. In Color based system we used histogram distance for observation model. In the integrated framework of particle filtering, we presented a new tracking algorithm based on instant / adaptive DT template matching by normalized cross-correlation. The template image is converted as DT image map before measuring similarity by normalized cross-correlation. The observation model is updated by the best matching score in every frame. Initialization is necessary to start the tracking process. In our proposed system model, we can easily initialize our system, by selecting any object in a video scene, and this selected object become template instantly adapt itself to the changing appearance or environments and thus make more robust tracking. The system has been tested on a variety of video data and very satisfactory results have been obtained. There are so many possibilities to extend our work such as for multiple object tracking, learning based tracking etc.

Acknowledgement

This research work was supported by MIC & IITA through IT leading R&D support project [2006-S-028-01].

References

[1] C. Yunqiang and R. Yong, "Real time object tracking in video sequences", *Signals and Communications Technologies, Interactive Video*, 2006, Part II: pp. 67 – 88.

[2] P. Li, T. Zhang, and A. E. C. Pece, "Visual contour tracking based on particle filters", *Image Vision Comput*., 2003, 21(1):111 – 123.

[3] V. T. Krishna and R. N. Kamesh, "Object tracking in video using particle filtering", In *Proc. IEEE Intl. Conf. on Accou., Speech, and Signal Proces,*. Volume 2, March 2005, pp. 657 – 660.

[4] A. Lehuger, P. Lechat and P. Perez, "An adaptive mixture color model for robust visual tracking", *In Proc. IEEE Int. Conf. on Image Process,* October, 2006, pp. 573 – 576.

[5] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "A color-based particle filter", In *Proc. First International Workshop on Generative-Model-Based Vision*, 2002, pp. 53 – 60.

[6] M. Israd and A. Blake, "CONDENSATION – conditional density propagation for visual tracking", *Int. journal of computer vision*, 1998, 29(1): 893 – 908.

[7] P. M. Djuric, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo and J. Miguez, "Particle Filtering*",* *IEEE signal processing magazine*, September 2003, pp. 19 – 38.

[8] B.D.R. Stenger, "Model-Based Hand Tracking Using A Hierarchical Baysian Filter", *Ph.D. Thesis,* University of Cambridge, March 2004.

[9] F. Zhao, Q. Huang and W. Gao, "Image Matching by Normalized Cross-Correlation", In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP2006.* May 2006, Volume 2, pp. II-II, 14-19.

[10] D.M. Gavrila, "Multi-feature Hierarchical Template Matching Using Distance Transforms", In *Proc. IEEE ICPR,* Brisbane, Australia, 1998.

[11] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching", In proc. *ICCV98*, January, 1998, pp. 107 – 112.

Authors

Md Zahidul Islam received his B.Sc. and M.Sc. degrees from the department of Applied Physics & Electronics Engineering, University of Rajshahi (RU), Bangladesh, in 2000 and 2002 respectively. In 2003, he joined as a lecturer in the department of Information Communication Engineering, Islamic University (IU), Kushtia, Bangladesh. With study leave from IU, in 2007, he is currently working on the development of visual object tracking system with various aspects as PhD candidate of Computer Engineering in Intelligent Image Media & Interface Lab, Chonnam National University (CNU), South Korea. His other current research interests include computer vision, 3D object, human and motion tracking and tracking articulated body.

Chi-Min Oh received the BS degree in computer engineering from Chonnam National University, Gwang-ju, Korea in 2007. Since February 2007, he has been pursuing the MS degree in Intelligent Image Media and Interface Laboratory in Chonnam National University. His research interests include Understanding Human Motions and Tracking Articulated Objects.

Chil-Woo LEE received the B.S. and M.S. degrees in electronic engineering from Chung-Ang University in 1986 and 1988 respectively in Seoul, Korea. And he received Ph.D. also in electronic engineering in 1992 from University of Tokyo, Japan. Since 1996, he has been Professor, Dept. of Computer Engineering, Chonnam National University in Gwangju, Korea. He has been worked as senior researcher at Laboratories of Image Information Science and Technology (LIST) for four years, form 1992 to 1996, and at that time he had an extra post of visiting researcher at Osaka University in Osaka, Japan. From January, 2001, he has visited North Carolina A&T University as a visiting researcher and jointly worked on several digital signal processing projects. Up to now, his research work has been associated with image recognition and image synthesis. His interests include Computer Vision, Computer Graphics, and visual human interface system. And he is also very interested in realization of real-time sensor system that can be aware of context of circumference.