Particle Filter for Human Positioning

Tassaphan Suwannatat

Mechatronics Engineering, Faculty of Science and Technology

Rajamangala University of Technology Tawan-ok, Bangphra Campus, Chonburi

E-mail: tassaphan_su@rmutto.ac.th Tel.0819409272

Abstract

Tracking moving human from a monocular vision using visual tracking is a useful skill for the coming generation of human-machine interface. It is a challenging problem for planning and controlling in dynamic environment. The methods used in existing moving human tracking that operated from fixed platform, fixed background, were not applicable. In this paper, visual approach detecting unexpected moving human that appears in the scene of a monocular camera was proposed. The system was designed to track and determine the location of the human with the use of conventional CCD camera. This method was implemented using algorithm based on the particle filter, which allowed accurate visual tracking under the circumstance of real-time visual tracking. A particle filter implementation to human tracking using human's model observation that exploits head-and-shoulder boundary as its prior density was applied. In our experiment, the results showed that the relevance of our approach to the problem of human tracking was then investigated. The accuracy was evaluated using the real visual tracking experiments. As a result, it was possible to track human's motion in the complex environment.

1. Introduction

Recently, cameras are cheap and ubiquitous sensors. Images provide much data about the environment. Therefore, the most widely used approach for tracking human motions is visual tracking. The existing techniques mainly differ in the features or models they use to extract humans from images. For example, an approach that relies on a fixed background to locate the human, the background image is subtracted to obtain a contour of the person (McHugh, 2009), (Kim, 2008), (Ahn, 2007). Other approaches rely on color for tracking (Guo, 2008), (Mohamed, 2007), (Quan, 2002). Additionally, there is a variety of vision-based techniques for tracking human. Existing approaches use stereo vision to identify humans (Bota, 2009), (Schmudderich, 2008), (Alonso, 2007). However, they do not apply any filtering techniques to keep track of the persons, so that they cannot reliably keep track of individual objects over time and they do not deal with other detection failures.

In the visual application, tracking is the problem of following an object in a sequence of image. The position and the shape of an object can be measured by several image processing algorithms (Zhang, 2008), (Raskin, 2007). These measures, however, do not always provide the true values of a moving object but can have high inaccuracies. The proposed tracking algorithms is to filter the unprecise measurements and to maintain an estimation of the current of an object. The state of an object typically consists of kinematic components like position, velocity and acceleration, and of additional features like the size and the shape of an object.

Unfortunately, the problem of tracking the position of moving human is significantly hard because human shape and size is flexible uniform (Zhang, 2008), (Raskin, 2007). The tracking process is harder since observations may result in ambiguities, features may not be distinguishable, objects may be occluded, or there may be more features than objects. In this paper we present a probabilistic framework for tracking a moving human. This technique uses a single camera and a motion model of the object being tracked in order to estimate their position and velocity.

This paper is organized as follows. After introducing the general concept of particle filter in the next section, we introduce the learning of human's model in Section 2. In Section 3, we present the particle filter and the implementation. Section 4 describes the experiments. Finally, Section 5 summaries the paper and makes the suggestions for future research.

2. Learning Human's Model

An essential ingredient in the proposed method is a prescription for choosing posterior density. This will require a prior density which could be specified by the user but for our application was learnt automatically by the following method.

2.1 A Shape Vector of Head-and-Shoulder Area

In order to derive a model to represent the shape of humans as they appear in the image, we represent it by vector sets of head-and-shoulder area of human's model. This is reasonable for the scenes we are dealing with where the people always appear in an upright position.

The system first segments moving human from a sequence by simple differencing from a background image and thresholding. The background image is subtracted from each input image of the sequence. The result image is thresholded to produce binary images in which the region in binary image corresponds to a moving human. After the region of a human is detected, the points around head-and-shoulder area are chosen. This vector set is reordered and stored as Q_i (shown in Fig. 1). It is important that the points are placed correctly on both skin-color and head-and-shoulder boundary.

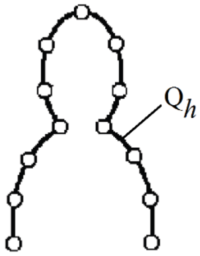


Fig.1 Human's mode

2.2 Aligning the Shape Vectors

Our method of aligning shape works by examining the statistics of the coordinates of the labeled points over the shape vector set. We achieve the required alignment by scaling, rotating and translating the training shapes. We aim to minimize a weighed sum of squares of distances between equivalent points on different shapes.

We first consider aligning a pair of shapes. Let Q_i be a vector describing the n + 1 points of i^{th} training set.

$$\mathrm{Q}_i = (x_{i0}, x_{i1}, \dots, x_{in}, y_{i0}, y_{i1}, \dots, y_{in})^{\mathrm{T}} \tag{1}$$

Let $M_j[\mathrm{Q}_j]$ be a rotation by j and a scaling by s_j. Given two similar shapes, Q_i and Q_j we can choose s_j and a translation $(t_x, t_y)_j$ mapping Q_i onto $M_j[\mathrm{Q}_j]$ so as to minimize the weighted sum.

where

$$M_j \begin{pmatrix} x_{jk} \\ y_{jk} \end{pmatrix} = \begin{pmatrix} (s_j \cos\theta_j)x_{jk} - (s_j \sin\theta_j)y_{jk} + t_{jx} \\ (s_j \sin\theta_j)x_{jk} - (s_j \cos\theta_j)y_{jk} + t_{jy} \end{pmatrix} \tag{2}$$

In order to aligning all the shapes in a set we use the following algorithm.

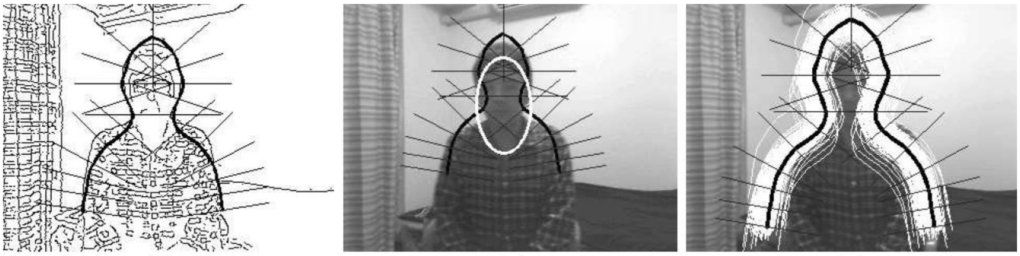
1. Select on shape to be the approximate mean shape.
2. Align the shapes to the approximate mean shape.
   - Calculate the centroid of each shape.
   - Align shapes to the origin.
   - Rotate, scale and translate the mean to the mean shape.
3. Calculate the new approximate mean from the aligned shapes
4. If the approximate mean from step 2 and 3 are different then return to step 2, otherwise return the true mean shape of the set.

$$\hat{\mathrm{Q}}_i = (\hat{x}_{i0}, \hat{x}_{i1}, \dots, \hat{x}_{in}, \hat{y}_{i0}, \hat{y}_{i1}, \dots, \hat{y}_{in})^{\mathrm{T}} \tag{3}$$

3. Particle Filter of Human's Model

Our proposed multi-features particle filter consists of 5 steps: selection, prediction, observation, measurement and report. The selection step is common to all particle filters. The process of our tracking algorithm is described as follows.

The process begins with a prior density and the effective prior for time-stamp $P(\mathrm{x}_t|\mathrm{z}_{t-1})$. It is derived from the output from the previous time-stamped sample set $P(\mathrm{x}_{t-1}|\mathrm{z}_{t-1})$, denote $s_{t-1}^{(n)}$ with weight $\pi_{t-1}^{(n)}$ and cumulative distribution $c_{t-1}^{(n)}$ for n=1, 2, ..., *N*. The first operation is to sample *N* times from the set $s_{t-1}^{(n)}$, choosing a given element with probability $\pi_{t-1}^{(n)}$. Each element chosen from the new set is now subjected to a predictive step. This corresponds to sampling from the distribution $P(\mathrm{x}_t|\mathrm{z}_{t-1})$. At this step, the sample set $s_t^{(n)}$ for the new time-stamp has been generated but, as yet, without its weight. It is approximately a fair random sample from the effective prior density $P(\mathrm{x}_t|\mathrm{z}_{t-1})$ from time-stamp *t*. Next, our developed observation step is applied, generating weights $P(\mathrm{x}_t|\mathrm{z}_t)$ from the predicted sampling $s_t^{(n)}$ and a new observation u_t to obtain the weights representation $\beta_t^{(n)}$ and α_t . These weighted sample-sets are used to measure the likelihood ratio of sample discriminant D_t at the measurement step. If D_t is higher than 1, then the sample set $s_t^{(n)}$ is set as the state of new observation $\mathbf{u}_t$ and $\pi_{t-1}^{(n)} = 1$ for 1, 2, …., N otherwise, $\pi_t^{(n)} = \beta_t^{(n)}$. Finally, after any time-stamp, it is possible to report on the current state by evaluating the moment of the state density.



(a) (b) (c)

Fig.2 (a) A result of shape space fitting, (b) A sample output of the observation from a new sample and (c) The mean of state density.

4. Experiments and Results

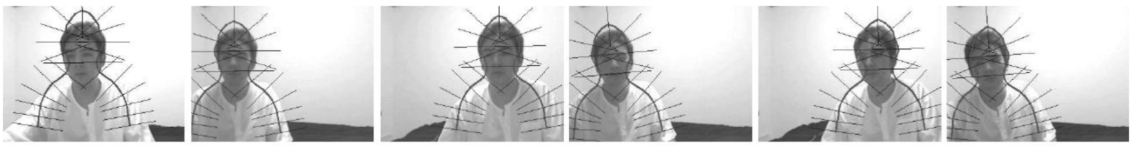
4.1 Testing of tracking parameters

In this section, we want to test how the filter parameters affect the tracking results. We examined the affectivity through the experiments of tracking a single human moved in the 3D space with constant velocity, while the human was tracked on the 2D image plane. Real video images were taken by a digital video camera with a resolution of 320 x 280 pixels. In the image sequence as shown in Fig.3, a person moves left and right with approximate velocity of human walking. The sequence had 719 frames with duration of 25s. The *shape space* template for tracking is built from the automatic learning as explained in Sect. 2.

Here, the accuracies of tracking are examined when dynamic noise parameters, number of particles and number of shape control points are varied. The accuracies are based on the mean tracking error which is defined as

$$\text{Mean Tracking Error} = \frac{1}{T}\sum_{i=1}^{T}|\mathbf{x}_i - \mathbf{r}_i| \quad (4)$$

where T is the total frame number of video sequence, x_i is the expected human state from filter estimation at time-stamp t. For comparison, we implemented r_i as the true human state at time-stamp t which is detected from background subtraction method.



(a) (b) (c) (d) (e) (f)

Fig. 3. Sub-sample images of the tracking parameter testing. (a) Frame 1, (b) Frame 124, (c) Frame 324, (d) Frame 426, (e) Frame 527 and (f) Frame 659.

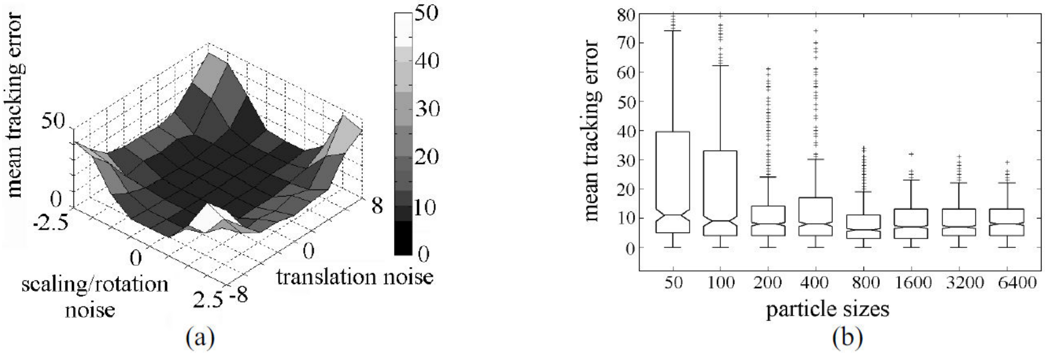


Fig. 4. (a) Mean tracking error of dynamic parameter testing and (b) Mean tracking error of particles sizes testing.

Dynamic Noise

Here, we examined the affectivity of dynamic noise parameters to the tracking results. In this experiment, the translation noise is varied from -8 to 8 and scaling/rotation noise is varied from -2.5 to 2.5. Each experiment, we used the number of particles of 41000 and used a trained model with 23 control points. The result of mean tracking error is shown in Fig. 4(a), the result had a region of high accuracy (low tracking error) at diffusing translation noise with magnitude ±4 and scaling/rotation noise between ±0.5. The accuracy strongly decreased when both translation and scaling/rotation noise were large, thus this tracking method is sensitive to the large magnitude of both translation and scaling/rotation noise.

Size of Particles

In this experiment, the accuracy of our method is examined by varying the number of particles from 50 to 6400 with a double step size. We set the translation noise as 2.5, scaling/rotation noise as 0.0005 and used a trained model with 23 control points. The dependence of the tracking error on number of particles is shown in Fig. 4(b). The performances of our tracker with particles number higher than 200 were better than those with a smaller number of particles.

Number of Control Points

This experiment examined the performance of tracker while numbers of control points of trained model varied from 3, 5, 11, 23, 45 and 89 points. As in Sect.3, we set translation noise as 2.5, scaling/rotation noise as 0.0005 and used the number of particles of 1000. Figure 5(a)-(f) show the output images at frame 249 of each experiment. The independence of the accuracy on the number of control points is shown in Fig. 5(g). The mean tracking errors are quite low for all the number control points (lower than 20 pixels) which mean that the accuracy of our tracking method is independent to the number of control points. These results suggest that having a large number of control points of the trained model does not directly guarantee a good performance of tracker.

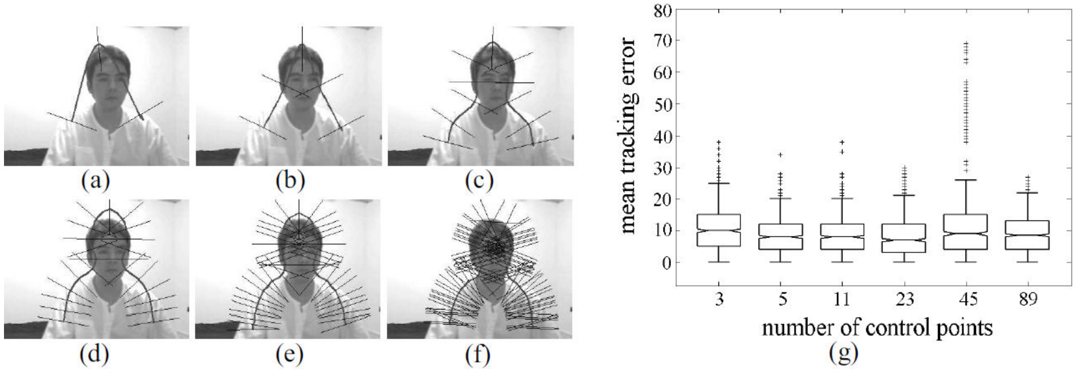


Fig. 5 Experiments on the various number of control points. (a) 3 points, (b) 5 points, (c) 11 points, (d) 23 points (e) 45 points (f) 89 points and (g) The results of mean tracking error.

5. Conclusions and Future work

In this paper, we proposed the particle filtering for human's positioning. Our proposed method emphasizes the detection of target likelihood that used the measurement of samples discriminant, a likelihood ratio which is derived from a probability model of current observation and prediction density. The method requires no previous knowledge of the background, is not affected by lighting changes, and effective in cluttered scenes. We also proposed the model of human as a priori distribution for the tracking by using contour feature-based of a human. Impressive results have been demonstrated in the experiments. In addition, our tracking method was examined through a real-time tracking of a human in front of a cluttered background through a moving camera. The results demonstrated that our method can track a position of a human throughout the sequence without lost tracking. To apply our method in various problems, however, improvement of the observation model should be based on the tasks. For example, the problem will become challenging when the number of targets is unknown and variable. An improvement in the observation model is one of the most important issues for our future work.

References

Ahn, J. and H. Byun. 2007. **Human Silhouette Extraction Method Using Region Based Background Subtraction**. Proceeding of LNCS, Computer Vision/Computer Graphics Collaboration Techniques, Vol. 4418. 412-420.

Alonso, I.P., D.F. Llorca, M.A. Sotelo, L.M. Bergasa, P.R. Toro, J. Nuevo, M. Ocana and M.A.G. Garrido. 2007. **Combination of Feature Extraction Methods for SVM Pedestrian Detection**. IEEE Transactions on Intelligent Transportation Systems, Vol.8, Issue 2. 292-307.

Bota, S., S. Nedevschi and Tomiuc, C. 2009. **Stereo-Based Pedestrian Detection for Collision-Avoidance Applications**. IEEE Transactions on Intelligent Transportation Systems, Vol.10, Issue 1. 1-13.

Guo, L. and J. Zhao. 2008. **Specific Human Detection from Surveillance Video Based on Color Invariant Moments**. Proceeding of IEEE Conference on the Second International Symposium on Intelligent Information Technology Application, (IITA'08), Vol.2. 331-335.

Kim, Z. 2008. **Real time object tracking based on dynamic feature grouping with back-ground subtraction**. Proceeding of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'08). 1-8.

McHugh, J., J. Konrad, V. Saligrama and Jodoin, P.M. 2009. **Foreground-Adaptive Background Subtraction**. Journal of IEEE on Signal Processing Letters, Vol.16, Issue 5. 390-393.

Mohamed, A.S.S., Y. Weng, Ipson S.S. J. and Jiang. 2007. **Face detection based on skin color in image by neural networks**. International Conference on Intelligent and Advanced Systems (ICIAS'07). 779-783.

Quan, H. M. Mitsuhiko and Masahide, K. 2002. **Skin-color-based image segmentation and its application in face detection**. IAPRWorkshop on Machine Vision Applications. 48-51.

Raskin, L., E. Rivlin and Rudzsky, M. 2007. **Using Gaussian Processes for Human Tracking and Action Classification**. Journal of LNCS, Advances in Visual Computing, Vol.4841. 36-45.

Schmudderich, J. V. Willert, J. Eggert, S. Rebhan, C. Goerick, G. Sagerer and E. Korner. 2008. **Estimating Object Proper Motion Using Optical Flow, Kinematics, and Depth Information**. IEEE Transaction on Systems, Man, and Cybernetics, Part B: Cybernetics, Vol.38, Issue 4. 1139-1151.

Zhang, Z. H. Gunes. and Piccardi, M. 2008. **Tracking People in Crowded by a Part Matching Approach**. Proceeding of IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance, (AVSS'08). 1926-1929.