A New Approach for Multiple Object Tracking – Discrete Event based Multiple Object Tracking (DEMOT)

Chi-Ho Kim*,**, Bum-Jae You**, Hagbae Kim*, and Sang-Rok Oh**

* Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Korea

**Intelligent Robotics Research Center, Korea Institute of Science and Technology, Seoul, Korea

(Tel : +82-2-958-5708; E-mail: chkim@yonsei.ac.kr)

(Tel: +82-2-958-5760; E-mail: ybj@amadeus.kist.re.kr)

Abstract: Tracking is a fundamental technique which is able to be applied to gesture recognition, visual surveillance, tangible agent and so forth. Especially, multiple object tracking has been extensively studied in recent years in order to perform many and more complicated tasks. In this paper, we propose a new approach of multiple object tracking which is based on discrete event. We call this system the DEMOT (Discrete Event based Multiple Object Tracking). This approach is based on the fact that a multiple object tracking can have just four situations - initiation, continuation, termination, and overlapping. Here, initiation, continuation, termination, and overlapping constitute a primary event set and this is based on the change of the number of extracted objects between a previous frame and a current frame. This system reduces computational costs and holds down the identity of all targets. We make experiments for this system with respect to the number of targets, each event, and processing period. We describe experimental results that show the successful multiple object tracking by using our approach.

Keywords: tracking, multiple object tracking, discrete event, initiation event, continuation event, termination event, overlapping event, discrete event based multiple object tracking (DEMOT)

1. INTRODUCTION

Tracking is a fundamental technique which is able to be applied to gesture recognition [1], visual surveillance [2], ubiquitous vision system [3], Human-Computer Interaction (HCI) [4], tangible agent, and so forth. However, most of these applications consider only one object in the scene. In order to perform many and more complicated tasks, the systems should consider multiple objects. Therefore, although multiple object tracking results in numerous tracking problems such as associating measurements with the inappropriate tracks and occlusion of targets that yields unacceptable tracking, it has been extensively studied in recent years.

Multiple object tracking has application in both military and civilian areas. For example, application areas include ballistic missile defense, air defense, ocean surveillance, and battle field surveillance. And, application areas include traffic control, home assistant robot, video surveillance and monitoring in banks, stores and parking lots [5], and camera motion and/or 3-D scene geometry estimation, too. Typically, predictions are first made as to the expected locations of targets in a current frame. These expected locations are then matched to actual measurements. At this stage, ambiguities may arise. Predictions for the expected location may not be supported by measurements – have these objects terminated or were they simply overlapped? There may be unexpected measurements – do these measurements originate from new tracks or are they spurious readings from noises? More than one measurement may match a predicted track – which measurement is the correct one and what is the origin of the other measurements? Or a single measurement may match to more than one track – which track should the measurement be assigned to? These ambiguities must be resolved in order to solve the motion correspondence problem.

There are many methods of multiple object tracking. However, almost all methods has assumed that the motion correspondence problem has been solved or is trivial so that a nearest neighbor strategy is effective. In some cases, a nearest neighbor strategy is indeed adequate. For example, Tomasi and Kanade [6] track corner features over very many frames using such an approach. A nearest neighbor strategy usually relies on the frame-to-frame image motion being extremely small. However, much more data must be processed. Moreover, if significant frame-to-frame image motions are present, ambiguities can quickly arise. Zheng and Chellappa [7] minimize these ambiguities by using a weighted correlation window to detected tracks in the next frame. Correlation techniques can significantly reduce the motion correspondence ambiguity. However, partial occlusion and significant changes in the background can be problematic for such methods. Moreover, such techniques are only appropriate to the detection of measurements from existing tracks, not for the detection of new tracks. Many researchers have used the Kalman filter to track geometric features such as lines [8] and corners [9] in as scene, under the assumption that motion correspondence is straightforward. The motivation and significance of this work is in designing stable and reliable algorithms to infer the 3-D structure and motion from 2-D image plane measurements. Shapiro, Wang, and Brady [10] describe tracking corners in the image plane. However, the motion correspondence problem is not rigorously addressed. Thus the target tracking and surveillance community has extensively studied the motion correspondence problem [11] and a number of statistical data association techniques have been developed. These algorithms are now receiving wider attention, especially within the computer vision community [12]. For example, Chang and Aggarwal [13] have applied the joint probabilistic data association (JPDA) filter [14] to the problem of 3-D structure reconstruction from an ego motion sequence. However, the JPDA is appropriate only if the number of tracks is known a priori and remains fixed throughout the motion sequence. Zhang and Faugeras [15] have used the track splitting filter of Smith and Buechler [16] for dynamic motion analysis. However, the track splitting filter allows measurements to be shared between tracks. This is physically unrealistic. More reasonable, is that a measurement originates from only a single source feature. Thus the multiple hypotheses tracking (MHT) originally proposed by Reid [17] is used for multiple object tracking. The foremost difficulty in the application of multiple object tracking involves the problem of associating measurements with the appropriate tracks, especially when there are missing reports (probability of detection less than unity), unknown targets (requiring track initiation), and false reports (from clutter). The key development of the MHT is a method for calculating the probabilities of various data-association hypotheses. With this development, the synthesis of a number of other features becomes possible. However, it becomes computationally expensive both in time and memory as the number of measurements increases. Thus there are many efforts in order to limit the number of hypotheses. Cox and Hingorani [18] describe an efficient implementation of the MHT algorithm and evaluate its usefulness in the context of visual tracking and motion correspondence. And Polat, Yeasin, and Sharma [19] incorporate a path coherence function along with MHT to reduce the negative effects of spurious measurements that produce unconvincing tracks and needless computations. However they have still a burden for computational costs. Kuno [2] proposes a multiple object tracking approach using the optical flow in order to track objects in a complex scenes such as outdoors where there are various noises such as changes of lighting and movements of background objects. However, the optical flow itself also needs many computational quantities.

In this paper, we propose a new approach for multiple object tracking which is based on discrete event. We call this system the DEMOT (Discrete Event based Multiple Object Tracking). Here, initiation, continuation, termination, and overlapping event constitute a primary event set and are fired by reasonable conditions. We extract features by color and motion segmentation together. Here, we use the smooth color model robust to irregular illumination condition that we have developed [20]. And we use the line tracker [21] that can find the exact profile of objects and perform real-time processing. And we use the concept of *focus of attention* algorithms [22, 23] to monitor a new track. This also contributes to real-time processing. Moreover, the DEMOT constituted by these methods holds down the identity of all tracks.

Section 2 introduces the DEMOT and constructs its framework. Subsequently in section 3 we describe the visual tracking techniques for the construction of the DEMOT – methods for feature extraction, line tracker, and focus of attention. Finally in section 4, we present the experimental results and section 5 concludes this paper. Especially, it is verified that the performance of the DEMOT is excellent in the viewpoint of real-time multiple object tracking from experimental results.

2. DISCRETE EVENT BASED MULTIPLE OBJECT TRACKING (DEMOT)

A Discrete Event System (DES) is a discrete-state, event-driven system, that is, its state evolution depends entirely on the occurrence of asynchronous discrete events over time [24]. Our approach for a multiple object tracking - DEMOT - is based on the concept of DES. A multiple object tracking can have just four situations of the following:

1) *Initiation*. The automatic creation of the new track when a new target enters the field of view.
2) *Continuation*. The continuation of a track through several frames.
3) *Termination*. The automatic termination of the track when a target is no longer visible for an extended period of time.
4) *Overlapping*. The automatic maintenance of the tracks when targets overlap.

The above four situations constitute the primary event set of DEMOT. The events are fired by reasonable conditions. Fig.1 is the Petri net representation of DEMOT system.

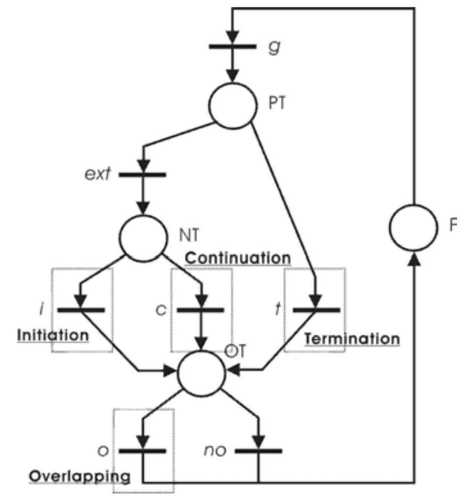


Fig.1 Petri net graph for DEMOT

The indices in Fig.1 mean the following:

▷ Transition set: T={*g*, *ext*, *i*, *c*, *t*, *o*, *no*}
- *g*: grabbing
- *ext*: There are all previous tracks.
- *i*: **Initiation** event
- *c*: **Continuation** event
- *t*: **Termination** event
- *o*: **Overlapping** event
- *no*: No overlapping track exists.

▷ Place set: P={PT, NT, OT, F}
- PT: Find previous tracks.
- NT: Find a new track.
- OT: Find overlapping tracks.
- F: Finish.

To illustrate the process of firing transitions and changing the state of a Petri net, consider the Petri net of Fig.2a, where the initial state is

$$\mathbf{x}_0 = [0,0,0,1].$$

We can see that the only transition enabled is *g*, since it requires a single token from place F. When *g* fires, one token is removed from F, and one token is placed in place PT, where the new state is

$$\mathbf{x}_1 = [1,0,0,0]$$

as shown in Fig.2b. This represents that this system grabs one frame and determines whether all previous tracks exist or not. With the same method, when *ext* fires, one token is removed from PT, and one token is placed in place NT, where the new state is

$$\mathbf{x}_2 = [0,1,0,0]$$

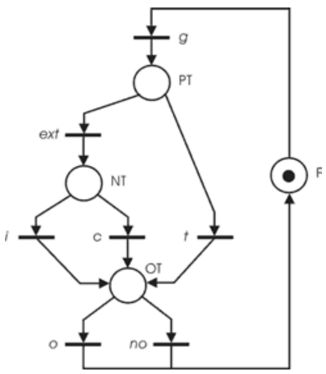
as shown in Fig.2c. This represents that there are all previous tracks and this system determines whether a new track exists or not. In this state, two transitions *i* and *c* are enable. Suppose transition *i* fires. One token is removed from place NT. The output place is OT. Therefore, a token is added to OT. The new state is

$$\mathbf{x}_3 = [0,0,1,0]$$

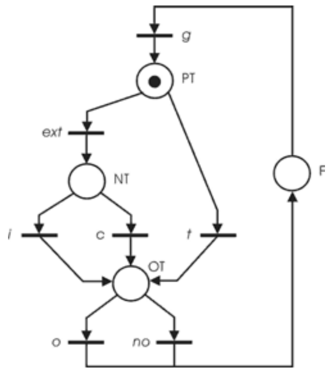
as shown in Fig.2d. This represent that this system performs initiation event and because of the existence of a new track determines whether overlapping tracks exist or not. In this state, two transitions *o* and *no* are enable, too. Suppose transition *o* fires. One token is removed from place OT. The output place is F. Therefore, a token is added to F. The new state is

$$\mathbf{x}_4 = [0,0,0,1]$$

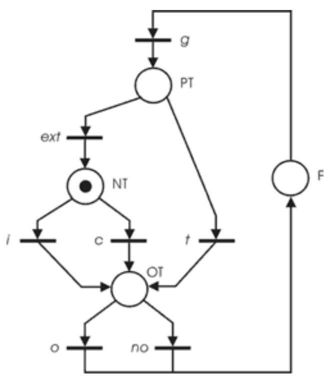
as shown in Fig.2a. This represent that this system performs overlapping event because of the existence of overlapping tracks and finishes the process for this frame.



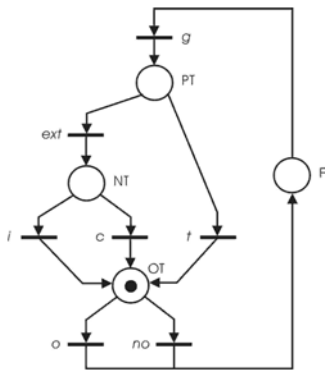
a. Initial state



b. New state resulting from firing transition *g*



c. New state transition from firing transition *ext*



d. New state transition from firing transition *i*

Fig.2 An example for the process of firing transition in the DEMOT

3. VISUAL TRACKING TECHNIQES FOR THE CONSTRUCTION OF THE DEMOT

We use the following three techniques that are primary methods for visual tracking. We perform the DEMOT by using these techniques. And we can hold down the identity of all targets by using these techniques.

3.1 Feature extraction

We use color and motion segmentation together to extract features. Here, we use the color segmentation algorithm robust to irregular illumination condition that we have developed [20]. This algorithm is based on the fact that a color has difficult statistical characteristics at different intensities. So, this color model results in modeling statistical characteristics of a color with respect to intensity. The statistical characteristics include the means and standard deviations of hue and saturation with respect to intensity. After a number of images in different illumination conditions are stored, the color modeling process for a single-colored target object starts by the following steps.

Step 1) Divide the total range of intensity into finite number of intervals, N. We use N=5.
Step 2) Collect pixels whose intensity belongs to the same intensity interval. The region for the target object is set manually.
Step 3) Make four bar graphs for the mean and standard deviation of hue and saturation with respect to intensity intervals.
Step 4) Approximate each bar graph as a continuous curve by using second-order B-spline curves. And, find the four functions, $H_m(x), H_\sigma(x), S_m(x)$, and $S_\sigma(x)$.

So, pixels that satisfy the following two conditions for hue and saturation are classified as a color.

$$H_m(x) - T_h \cdot H_\sigma(x) \le H(p(x)) \le H_m(x) + T_h \cdot H_\sigma(x)$$

$$S_m(x) - T_h \cdot S_\sigma(x) \le S(p(x)) \le S_m(x) + T_h \cdot S_\sigma(x)$$

Here,

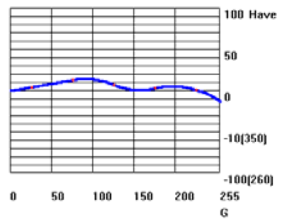
$H(p(x))$: Hue of $p(x)$,

$S(p(x))$: Saturation of $p(x)$,

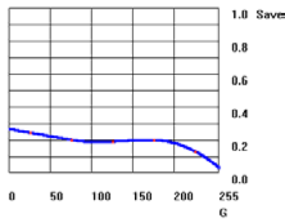
$p(x)$: Pixel with the intensity x, and

T_h : Threshold value.

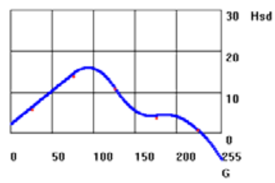
The model made for the DEMOT is presented by Fig.3. The curves of Fig.3 are the average of hue, the average of saturation, the standard deviation of hue, and the standard deviation of saturation from Fig.3a to Fig.3d, respectively. And, the horizontal axes are the value of intensity.



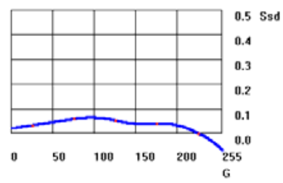
a. The average of hue



b. The average of saturation



c. The standard deviation of hue



d. The standard deviation of hue

Fig.3 The color model made for the DEMOT

The result of skin color clustering by using this color model is shown in Fig.4. Fig.4a is an input image acquired in the laboratory and Fig.4b is a result for that image. We verify that the skin area is successfully extracted. A little noise will be removed by motion segmentation.

a. Input image b. Resultant image

Fig.4 The results of skin color clustering by using this color model

3.2 Line tracker

The existing visual tracking methods mainly segmented the target region by using image processing methods such as region growing. Although these methods exactly define the profile of target, these are suited to real-time tracking for the so much computational costs. So, we use four line trackers - top tracker, bottom tracker, left tracker, and right tracker - for the real-time visual tracking. First, a tracker is moved with a fixed distance to the direction in which the number of pixels extracted by the above color and motion segmentation algorithm is not equal to 0. And if the number of pixels verified as a target is also not equal to 0, a tracker is moved with a fixed distance to the same direction. We repeat this process until the number of pixels verified as a target becomes 0. Here, the location in which the number of pixels verified as a target is equal to 0 is the location of the tracker for this target in a current frame. We determine the location of other three trackers with the same method. These four trackers that is connected each other define the region of the targets. We achieve the multiple object tracking by performing this process for the all targets. Especially, after tracking, if the top tracker is lower than the bottom tracker or the left tracker is more right than the right tracker in a target, this system regards that the target is disappeared and performs termination event.

The line tracker can not only realize a real-time visual tracking but also automatically solve motion correspondence problem.

3.2 Discriminative of Focus of Attention (DFA)

We propose another new method to realize real-time processing. We call it Discriminative Focus of Attention (DFA). We use DFA to monitor new track. We divide focus of attention into the following two levels:

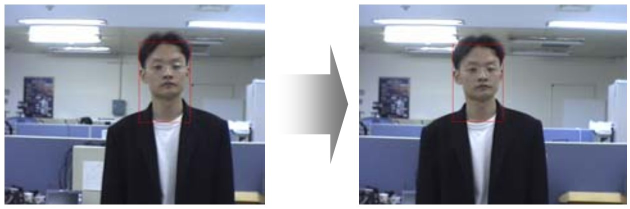
- *level 1*. We use this to monitor the tracks of a previous frame. This level is processed with high resolution.
- *level 2*. We use this to monitor a new track in a current frame. This level is processed with low resolution.

We use 10×10 super pixel as the resolution of level 2. DFA constitutes to real-time processing by decreasing the resolution of a background.

4. EXPERIMENTAL RESULTS

We make experiments for the proposed algorithm on a 2.4 GHz PC-based vision system equipped with Matrox CORONA frame grabber. Sony XC-999 color camera is used. We use body parts (face and hands) as multiple targets.

Fig.5 is the results of the DEMOT with respect to the number of targets. The number of targets is 1, 2, and 3 from Fig.5a to Fig.5c, respectively. Fig.5 shows that the DEMOT successfully operates irrespective of the number of targets.



a. The number of target is 1.

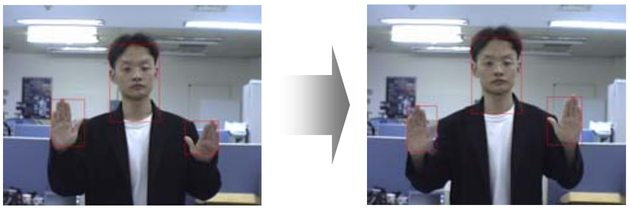
b. The number of targets is 2.

c. The number of targets is 3.

Fig.5 The results of the DEMOT with respect to the number of targets

The results of the DEMOT and the state transition with respect to each event are shown in Fig.6. Fig.6 shows the results of the DEMOT and the state transition with respect to the initiation, continuation, termination, and overlapping event from Fig.6a to Fig.6d, respectively. We verify that the DEMOT successfully operates with respect to its all four events.

[0,1,0,0]→[0,0,1,0]
a. Initiation event



[0,1,0,0]→[0,0,1,0]
b. Continuation event

[1,0,0,0]→[0,0,1,0]
c. Termination event

[0,0,1,0]→[0,0,0,1]
d. Overlapping event

Fig.6 The results of the DEMOT and the state transition with respect to each event

Table 1 shows processing period with respect to the number of targets. Period just a little increases as the number of targets increases. So, we confirm that the DEMOT has the property of real-time processing.

Table 1 Processing period with respect to the number of targetss

The number of target(s)	Period (msec/frame)
0	66.72
1	67.34
2	97.50
3	99.53

And this system successfully holds down the identity of targets.

Additionally, Fig.7 shows the results of the DEMOT with respect to various illumination conditions. We verify that the DEMOT successfully also operates in various illumination conditions because we use our smooth color model robust to brightness variation..

Fig.7 The results of the DEMOT with respect to various illumination conditions

5. CONCLUSION

We have proposed a new approach for multiple object tracking – Discrete Event based Multiple Object Tracking. We call this DEMOT. This approach is based on the fact that a multiple object tracking can have just four situation, that is, initiation, continuation, termination, and overlapping. Here, initiation, continuation, termination, and overlapping constitute a primary event set and this is based on the change of the number of object extracted in each frame. We have made experiments for the proposed algorithm in laboratory environment. First, we have experimented on the DEMOT with respect to the number of targets. We have verified from the results that the DEMOT successfully operates irrespective of the number of targets. And, we have experimented on the DEMOT with respect to its all four events. We have verified from the results that the DEMOT successfully operates with respect to each event. Finally, we have measured the processing period of the DEMOT with respect to the number of targets. We have verified from the results that the DEMOT has the property of real-time processing. And, It has been confirmed that the DEMOT have held down the identity of targets by using several visual tracking techniques for the construction of the DEMOT. We will use this system as the fundamental technique of many applications such as ubiquitous vision system, tangible agent, and so on.

REFERENCES

[1] J.Triesch, "Vision-based Robotic Gesture Recognition," *Dissertation for Getting the Degree of Doctor of Natural Science in the Faculty for Physics and Astronautics of the Ruhr-University of Bochum*, May 1999.
[2] Y.Kuno, "Detection and Tracking People in Complex Scenes," *P.Remagnio, G.A.Jones, N.Paragios, and C.S.Regazzoni, ed., Video-based Surveillance Systems:*

Computer Vision and Distributed Processing, Kluwer Academic Publishers, pp.65-78, 2002.
[3] K.C.Ng, H.Ishiguro, M.Trivedi, and T.Sogo, “Monitoring Dynamically Changing Environments by Ubiquitous Vision System,” *Second IEEE Workshop on Visual Surveillance*, pp.67-73, 26 June 1999.
[4] Y.Wu and T.S.Huang, “Nonstationary Color Tracking for Vision-based Human-Computer Interaction,” *IEEE Trans. Neural Networks*, vol.13, pp.948-960, July 2002.
[5] Video Surveillance and Monitoring Homepage (http://www-2.cs.cmu.edu/~vsam)
[6] C.Tomasi and T.Kanade, “Shape and Motion from Image Streams under Orthography: A Factorization Method,” *Int’l J. of Computer Vision*, vol.9, no.2, pp.137-154, 1992.
[7] Q.Zheng and R.Chellappa, “Automatic feature Point Extraction and Tracking in Image Sequences for Unknown Camera Motion,” *Proc. Fourth Int’l Conf. on Computer Vision*, pp.335-339, 1993.
[8] N.Ayache and O.Faugeras, “Maintaining Representations of the Environment of a Mobile Robot,” *IEEE Trans. on Robotics and Automation*, vol.5, no.6, pp.804-819, 1989.
[9] T.J.Broida, S.Chandrashekhar, and R.Chellappa, “Recursive 3-D Motion Estimation from a Monocular Image Sequence,” *IEEE Trans. on Aerospace and Electronic Systems*, vol.26, no.4, pp.639-656, 1990.
[10] L.S.Shapiro, H.Wang, and J.M.Brady, “A Matching and Tracking Strategy for Independently Moving Objects,” Proc. British Machine Vision Conf., pp.306-315, 1992.
[11] Y.Bar-Shalom and T.E.Fortmann, *Tracking and Data Association*, Academic Press, 1988.
[12] I.J.Cox, “A Review of Statistical Data Association Techniques for Motion Correspondence,” *Int’l J. of Computer Vision*, vol.10, no.1, pp.53-66, 1993.
[13] Y.L.Chang and J.K.Aggarwal, “3D Structure Reconstruction from an Ego Motion Sequence using Statistical Estimation and Detection Theory,” *Proc. IEEE Workshop on Visual Motion*, pp.268-273, 1991.
[14] T.E.Fortmann, Y.Bar-Shalom, and M.Scheffe, “Sonar Tracking of Multiple Targets using Joint Probabilistic Data Association,” *IEEE J. of Oceanic Eng.*, vol.8, no.3, pp.173-184, 1983.
[15] Z.Zhang and O.D.Faugeras, “Three-dimensional Motion Computation and Object Segmentation in a Long Sequence of Stereo Frames,” *Int’l J. of Computer Vision*, vol.7, no.3, pp.211-241, 1992.
[16] P.Smith and G.Buechler, “A Branching Algorithm for Discriminating and Tracking Multiple Objects,” *IEEE Trans. on Automatic Control*, vol.20, pp.101-104, 1975.
[17] D.B.Reid, “An Algorithm for Tracking Multiple Targets,” *IEEE Trans. on Automatic Control*, vol.24, no.6, pp.843-854, Dec. 1979.
[18] I.J.Cox and S.L.Hingorani, “An Efficient Implementation of Reid’s Multiple Hypothesis Tracking Algorithm and Its Evaluation for the Purpose of Visual Tracking,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.18, no.2, Feb. 1996.
[19] E.Polat, M.Yeasin, and R.Sharma, “Tracking Body Parts of Multiple People: A New Approach,” *Proc. IEEE Workshop on Multi-object Tracking*, pp.35-42, 2001.
[20] C.H.Kim, B.J.You, H.Kim, and S.R.Oh, “A Robust Color Segmentation Technique for Tangible Space,” *15th Triennial Congress Int’l Ergonomics Association*, Aug. 24-29, Seoul, Korea, 2003. (to be published)
[21] Y.B.Lee, B.J.You, S.W.Lee, and K.B.Kim, “Real-time Face Tracking Algorithm Robust to Illumination Variations,” *Proc. KIEE Summer Annual Conference*, July 17-20, 2003.
[22] J.K.Tsotsos, “Towards a Computational Model of Visual Attention,” T.Papathomas, C.Chubb, A.Gorea, and E.Kowler, ed., *Early Vision and Beyond*, MIT Press, 1995.
[23] K.Toyama and G.D.Hager, “Incremental Focus of Attention for Robust Visual Tracking,” *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.189-195, June 18-20, 1996.
[24] C.G.Cassandras, *Discrete Event System: Modeling and Performance Analysis*, IRWIN, 1993.