Estimating the Positions and Postures of Non-rigid Objects lacking Sufficient Features based on the Stick and Ellipse Model

Norimichi Ukita[†], Toshihiro Kitajima and Masatsugu Kidode
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma, Nara, JAPAN
{ukita,kidode}@is.naist.jp

Abstract

We propose a method for tracking non-rigid objects, using an object model generated automatically from a set of sample images. Our model consists of multiple sticks and ellipses which represent the skeleton and the areas of an object, respectively. Because appearance features have to be extracted, previous methods cannot estimate the whole area and posture for 2-D image of a non-rigid object lacking sufficient characteristic features (e.g., texture patterns, shape and so on) to be detected easily. With the proposed model, on the other hand, our method can work well because (1) each component of the model can fit each rigid part of a non-rigid object and (2) the reliability of each component is evaluated. To confirm the effectiveness of the proposed method, we conducted several experiments with goldfish. The tracking system automatically generated a model of the goldfish, and could then track goldfish even when they were partially occluded.

1. Introduction

To understand dynamic situations in the real world, object detection and tracking is one of the most fundamental technologies. We can obtain information about observed objects (e.g., their number, velocity, and locus) with an object detection and tracking method. Moreover, by estimating the posture of the object, more detailed information (e.g., behavior and activity characteristics) can be acquired and the tracking method can then be employed in various applications.

When multiple objects exist in an observed scene, mutual occlusion may occur between them and interfere with continuous tracking. Many studies have been reported to solve this problem. In particular, several methods[1, 2, 3] that incorporate stochastic dynamics into the probabilistic framework have recently gained wide attention. While these methods can cope with occlusions, and can be widely applied, they have the following problems:

Problem 1 Cannot estimate the posture of a non-rigid object (defined as an articulated object whose joint positions are unknown):

In previous methods, the area of an object is represented as a simple configuration (e.g., an ellipse or a rectangle). Therefore, only a simple rigid object or a part (or several parts) of a complex non-rigid object can be tracked. For example, the methods proposed in [4] and [3] realized tracking of multiple cars (i.e., rigid objects) and human heads (i.e., a part of a non-rigid object), respectively.

To extract the whole region of each non-rigid object, on the other hand, an active contour model is effective. For example, [5] and [6] employ the Snake model and the Level set method, respectively. However, these methods have the following problems; (1) If a region of an object split into two or more regions due to occlusion, these regions are regarded as different objects after the breakup and (2) a silhouette of an object is extracted, but its posture cannot be estimated.

Problem 2 Cannot track the whole area of an object which has insufficient features:

In [7], multiple moving people can be tracked by employing an appearance-based human body model consisting of textual and shape components. In [8], an arbitrary non-rigid object can be tracked based on the mean shift iterations and the method can handle in real-time partial occlusions, significant clutter and scale variation. These methods, however, track the area of interest by extracting characteristic texture patterns, colors, shapes, and so on (e.g., skin color and the head of a human subject). It is difficult to find characteristic features

[†]Precursory Research for Embryonic Science and Technology, Japan Science and Technology Agency (PRESTO, JST)

in the whole of an arbitrary object, including objects with insufficient features such as animals and fish.

We solve the above problems as follows:

Solution 1 The proposed method divides the area of a target into multiple (semi)rigid parts, each of which is defined as a *part model*. All part models are integrated into an *object model*, where the end points of multiple part models are combined with each other and relative angles between them are variable. The geometric configuration of all the part models is considered to be the posture of the observed object.

Solution 2 To evaluate the likely area of an arbitrary object, the following information is estimated: skeleton (defined by the object's central line), color histogram, and rough shape (represented by an ellipse). These criteria are used to determine the number and positions of the part models, so that the whole area of the target is well represented and each part model can be detected easily. In addition, the adaptive reliability of each part model is evaluated to counter self and mutual occlusions.

Based on the above basic ideas, we represent the *object model* as the integration of the stick and ellipse models. The *stick model* corresponds to a straight skeleton that is a part of the center line of an object. The *ellipse model* contains information about the color and rough shape of an object. Our method can automatically generate an object model from a set of sample images, and the object model is used for tracking and estimating the positions and postures of multiple non-rigid objects.

2. Object model generation

This section describes the processing which generates the object model automatically from a set of sample images (video). In a set of sample images, only one object exists and it needs to be in various postures.

2.1. Stick model generation

It is possible to generate the stick model subjectively if the target is an articulated object whose joint positions are known (see [9, 10], for example). Since the target in this study is a non-rigid object whose joint positions are not known accurately, the system must estimate joint positions and the distances between joints from sample images. We assume that a bone in the skeleton can be approximated by a straight line; the system then extracts the skeleton from the target object's area by a thinning algorithm. The system represents rigid parts of the skeleton as sticks by the Hough transformation. Multiple sticks may be obtained from one target. The system generates the stick model by connecting multiple sticks and permitting changes of relative angles. The stick model in Fig. 2 was generated from the sample image in Fig. 1. Note that in a set of sample images, only one object exists and it needs to be in various postures. One stick model is generated from one sample image. Because the number of sticks and the connections between each stick change with the target's postures in sample images, the system selects the stick model derived from the most complicated posture, (that is, which has the most sticks in all sample images) to adjust the posture changes, and determines the selected model as the final stick model. Here, we call the stick whose center of gravity is nearest to that of the target area the standard stick, and each other stick has an identification number.

Figure 1. Sample image.

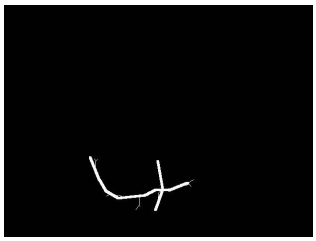

Figure 2. Stick model.

2.2. Stick reliability

The system cannot extract the skeleton corresponding to each stick in the stick model if the self occlusions occurs, because the skeleton corresponds to the central line of the target. Moreover, using the thinning algorithm, the system may model a stick which does not in fact exist due to the influence of noise. In such a case, the stick either does not correspond to the skeleton, or it corresponds to a skeleton derived from other portions of the target. If the system relies equally on all sticks in the stick model, it may cause tracking failure. In our method, the system coped with this problem by weighting the degree of stick reliability for every stick.

Here, the parameters of the standard stick are expressed by $\boldsymbol{\nu}^{std} = (x, y, \theta^{std})^{\top}$. (x, y) represents the position of the stick, and θ^{std} represents the angle. The parameters of other sticks are expressed only by the relative angle to the connecting stick, because each stick connection must be maintained. The number of sticks in the stick model is expressed by n; if

i identifies a stick, the parameters are expressed by $\theta^i(i = 1, 2, \ldots, n)(i \neq std)$. The parameters of the stick model are expressed by $\mathbf{\Phi} = \{\theta^1, \theta^2, \ldots, \nu^{std}, \ldots, \theta^n\}$.

To compute the reliability of each stick, the stick model as described in section 2.1 matches by changing the position and angle of the stick model, so that all sticks are within the target area. Matching degrees are the two following:

- The degree of overlap between the pixel on the target and on the stick:

The pixels on the target are expressed by $\boldsymbol{L}$; if i identifies a stick, the pixels on stick i are $\boldsymbol{S}^i$.

The degree of overlap of each stick is M^i_{ter}, and the degree of M_{ter} which represents the total of M^i_{ter} is computed by formula (1). Here, the number of all sticks in the stick model is n.

$$\begin{aligned} M^i_{ter}(\mathbf{\Phi}) &= N\left(\boldsymbol{S}^i(\mathbf{\Phi}) \cap \boldsymbol{L}\right) \,/\, N\left(\boldsymbol{S}^i(\mathbf{\Phi})\right) \\ M_{ter}(\mathbf{\Phi}) &= \sum_{i=1}^{n} M^i_{ter}(\mathbf{\Phi}), \end{aligned} \tag{1}$$

where $N(\boldsymbol{I})$ denotes the number of pixels included in $\boldsymbol{I}$.

- The degree of overlap between the pixels on the skeleton and on the stick:

The pixels on the skeleton of the target are expressed by $\boldsymbol{B}$. The degree of overlap of each stick is M^i_{skl}, and the degree of M_{skl} which represents the total of M^i_{skl} is computed by formula (2).

$$\begin{aligned} M^i_{skl}(\mathbf{\Phi}) &= N\left(\boldsymbol{S}^i(\mathbf{\Phi}) \cap \boldsymbol{B}\right) \,/\, N\left(\boldsymbol{S}^i(\mathbf{\Phi})\right) \\ M_{skl}(\mathbf{\Phi}) &= \sum_{i=1}^{n} M^i_{skl}(\mathbf{\Phi}) \end{aligned} \tag{2}$$

So that all sticks are in the target area, the parameters are limited by applying threshold processing to M_{ter}. The system determines the parameters having the highest degree of M_{skl} among the limited parameters as the parameters corresponding to the target area. The system judges whether the skeleton corresponding to stick i exists by applying threshold processing to M^i_{skl} in the determined parameters. This judgement is performed for every stick in the stick model. The same processing is performed for all sample images. The number of sample images in which skeletons corresponding to every stick exist is expressed by S^i_{ex}, and the total number of sample images is expressed as S_n.

Skeleton existence probability γ^i_{skl} is computed by $S^i_{ex} \,/\, S_n$. The reliability of each stick W^i_{stk} is computed by formula (3), so that the total of the reliabilities is 1.0.

$$W^i_{stk} = \gamma^i_{skl} \,/\, \sum_{i=1}^{n} \gamma^i_{skl} \tag{3}$$

2.3. Ellipse model generation

In our method, the ellipse model has rough shape and color histogram as information about the features representing the target. The shape information needs to correspond to the target outline exactly, because information which represents each partial area as exactly as possible is a more effective quantifier of the features. Then, the ellipse which is inscribed in the target outline is used as the part model, so that shape information represents the area as well as possible. Therefore, the ellipse expanding method is employed, which draws an inscribed circle in the outline and makes the long axis, the short axis, the angle, and the central point change to fit the outline. In our method, only the long and the short axes are changed when the circle expands. The central point is fixed on the stick, and the angle is fixed at the same angle as the stick, so that the area in the outline may be as accurate as possible.

The color histogram information needs to possess the characteristic colors so that the color histogram may become more effective as a feature. The system selects therefore a central point at which the ellipse includes the area of most characteristic colors.

First, the target area is expressed by the HSV table color system, and a hue histogram is created. Next, the system sequentially deletes hues of highest frequency, so that about 70 or 80 percent of the total hues are deleted. Then, rare colors, (that is, characteristic colors) can be detected from the target area. In the case of a goldfish, as shown in Fig. 3, the black portion of the eyes and the light red portion of the fins and tail are detected as the characteristic colors. It is necessary to determine the parameters of the stick model on the target area, because the central point of the ellipse is on the stick and the angle is the same angle as the stick. The parameters are then determined as described in section 2.2. Since any point can be the central point as long as it is a point on a stick, two or more candidates may exist. The ellipse is expanded in each candidate, and then the number of the pixels of the characteristic color in the ellipse is counted. The candidate which has the highest number of counts is determined as the central point of the ellipse, and the size of the ellipse is also determined from the long axis and the short axis. The ellipse model is completed by averaging the central point, the long axis, and the short axis in all sample images, because there is a possibility

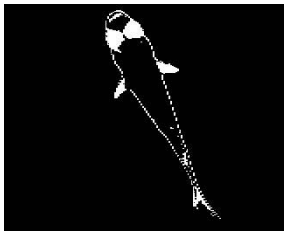

Figure 3. Area of the characteristic colors

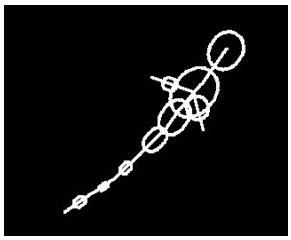

Figure 4. Object model.

that an ellipse model generated from only one sample image does not respond to various postures. The completed ellipse model is combined with the stick model, becoming the object model as shown in Fig. 4. Note that the hue histograms in the ellipses of all sample images are averaged and earned before tracking.

The reliability of the ellipse is expressed by Wc^i. Wc^i is based on the number of the pixels of the characteristic colors in the ellipse, because the ellipse including the area of most characteristic colors should correspond exactly to the target.

The color histogram information has to be learned before tracking. Therefore, the hue histograms in the ellipses of all sample images are averaged, and used as the learned histogram in tracking.

3. Tracking

In our method, the system tracks targets within the framework of the tracking method with a probability distribution [3]. The processing of tracking involves image capture, hypothesis generation, probability evaluation, probability distribution reform, probability distribution propagation, and then begins again with.

3.1. Probability evaluation

In order to track, the *probability* (which represents the possibility of the target's existence) must be determined as described in [3]. The probabilities of targets are independent of each other. Here, if i identifies a stick, the parameters of stick i are expressed by $\boldsymbol{P}^i = (x, y, \theta)^\top$. As mentioned later, unlike the parameter setup in 2.2, all sticks have three parameters for each position and angle in tracking (see 3.3). Probability is determined by the following three evaluations.

(a) Matching of the stick model and the skeleton:

In the direction vertical to the skeleton, the parameters are estimated with sufficient accuracy, even if the target is an object with insufficient features. The evaluation function E^i_{skl} by the stick model and the skeleton is:

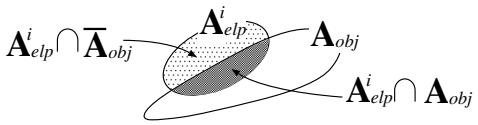

Figure 5. Area of an object and an ellipse.

$$E^i_{skl}(\boldsymbol{P}^i) = N\left(A^i_{stk}(\boldsymbol{P}^i) \cap A_{skl}\right) / N\left(A^i_{stk}(\boldsymbol{P}^i)\right) \quad (4)$$

Here, the area of stick i is A^i_{stk}, and the area of the skeleton is expressed by A_{skl}.

(b) Similarity of the hue histogram learned beforehand and that observed in tracking:

The position of the portion of the target which has the characteristic colors can be estimated with sufficient accuracy, using the color information. The function g^i_{clr} of the hue histogram is given by:

$$g^i_{clr}(\boldsymbol{P}^i) = \sum_{s=1}^{S} \min[\, I^i(s), M^i(s)\,] \quad (5)$$

Here, the number of hues in the histogram is S, the s-th hue ($s = 1, 2, \ldots, S$) in the histogram is s, the frequency of the hue histogram in the ellipse of stick i observed in tracking is I^i, and the frequency of the one learned beforehand is M^i. The total frequency of the hue histogram learned beforehand is ϕ^i_{clr}. The evaluation function E^i_{clr} in the hue histogram is computed by using g^i_{clr} with ϕ^i_{clr}. Then, the ellipse with the characteristic color becomes more reliable by using the reliability Wc^i of the ellipse computed in section 2.3. The evaluation function E^i_{clr} is:

$$E^i_{clr}(\boldsymbol{P}^i) = Wc^i \times \left(\, g^i_{clr}(\boldsymbol{P}^i) \,/\, \phi^i_{clr}\right) \quad (6)$$

(c) Matching of the area of the ellipse and the target:

If the configuration of each portion area of the target is different, information about matching of the area is more effective for estimating the target's positions and postures. The evaluation function E^i_{area} by matching of the area:

$$E^i_{area}(\boldsymbol{P}^i) = N\left(A^i_{elp}(\boldsymbol{P}^i) \cap A_{obj} - A^i_{elp}(\boldsymbol{P}^i) \cap \overline{A}_{obj}\right) / N\left(A^i_{elp}(\boldsymbol{P}^i)\right) \quad (7)$$

Here, the area of the ellipse of stick i is A^i_{elp}, and the area of the target is A_{obj}. As shown in Fig. 5, $A^i_{elp} \cap A_{obj}$ is the matching area, and $A^i_{elp} \cap \overline{A}_{obj}$ is the non-matching area in the ellipse.

The probability C^i in stick i is combined the three above-mentioned evaluation functions:

$$C^i(\boldsymbol{P}^i) = W_{skl}E^i_{skl}(\boldsymbol{P}^i) + W_{clr}E^i_{clr}(\boldsymbol{P}^i) + W_{area}E^i_{area}(\boldsymbol{P}^i) \quad (8)$$

where, W_{skl}, W_{clr}, and W_{area} are the weights of each evaluation function. The evaluation function of the whole object model is expressed by C. C is computed using each reliability given in section 2.2:

$$C(\boldsymbol{P}) = \sum_{i=1}^{n} W^i_{stk}C^i(\boldsymbol{P}^i) \quad (9)$$

where, $\boldsymbol{P}(\boldsymbol{P}^0, \boldsymbol{P}^1, \ldots, \boldsymbol{P}^n)$ are all the parameters of each stick.

3.2. Hypothesis generation

When the system starts tracking, it is necessary to estimate the target parameters from only the initial frame, because no prediction from a previous frame can be made. In the initial frame, parameters corresponding to the target are represented as a probability distribution in parameter space, which we call the hypothesis. The system tracks targets using hypotheses.

New objects may appear in the image after the initial frame. Then, since the area where hypotheses do not exist in any of the target areas is the area where new targets may appear, hypotheses should be generated in that region. We call the region where targets exist but are not initially apparent and the target area in the initial frame a *non-detected area*. In order to track a target in response to its emergence quickly, non-detected areas is found at each frame.

In non-detected areas, parameters corresponding to the targets have to be estimated. To do this, the method given in section 2.2 is employed to match the object model and the non-detected area. Parameters estimated by matching are expressed by Φ. If $M_{ter}(\Phi)$, defined by formula (1), is larger than a threshold, a new hypothesis is generated. Otherwise, the non-detected area is not regarded as a target region.

Although Φ are those parameters represented by three parameters in the standard stick and by one parameter in other sticks, all sticks can acquire three parameters by using the connection of the sticks for subsequent processing. These acquired parameters are expressed by P.

3.3. Probability distribution reform

Although the parameters $\boldsymbol{\Phi}$ represented by three parameters (x, y, θ) in the standard stick and by one parameter (θ) in other sticks are used when the object model is generated, the parameters $\boldsymbol{P}$ represented by three parameters in all sticks are used in tracking. The reason for not using the same parameter setup is that a probability distribution is used in tracking.

Only the standard stick's parameters can be treated independently in the parameter setup of $\boldsymbol{\Phi}$. Therefore, a multi-dimensional parameter space ($(3+n-1)$-dimension, where n is the stick number) is generated. If a probability distribution is generated in the parameter space, the parameters in the probability distribution of a particular stick will overlap each other; that is, their redundancy will be high. This leads to the problem of a waste of processing time.

On the other hand, since the parameters of each stick can be treated independently in the parameter setup of $\boldsymbol{P}$, the three-dimensional parameter space of the same number as the numbers of sticks is generated. In this case, the parameters in the probability distribution do not overlap each other, and redundancy is low. Since processing time is shortened, this parameter setup is suitable for real-time tracking. Moreover, the search range for every stick can be expanded and reduced, because variations of the probability distribution for every parameter space can be accommodated.

A propagated probability distribution does not reflect the information within a newly captured image. To accurately reflect this information, the propagated probability distribution has to be reformed through feedback from the new image. In earlier methods, parameters which had the maximum probability were selected as the center of the new probability distribution, and the probability distribution was reformed with the parameters. In our method, however, if the probability distribution is reformed using parameters which have the maximum probability as the center in each stick, it may be impossible to maintain the connectivity of sticks and the sticks will come apart, because the parameter setup $\boldsymbol{P}$ is used in tracking. Therefore, the probability distribution has to be reformed while maintaining stick connectivity.

We assume that stick i can be relied on, and it is called the reliance stick rt. The parameters which have the maximum probability in the probability distribution of rt are determined as the representative parameters of rt. Then, the parameters in the probability distribution of stick a, which connects with rt, are restricted, because it must connect with rt. As shown in Fig. 6, the parameters of rt and a can then be described by a tree structure with a root node and branch nodes $a1, a2, a3$ which are the restricted parameters of stick a. The node represents the probability of the parameters of each stick. Next, the parameters of stick b which connects with a are restricted in the same way: the parameters of a and b are described by node $a1$

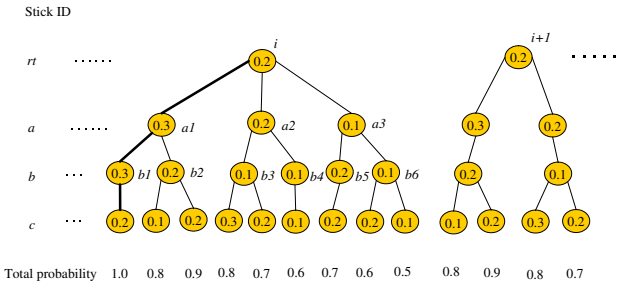

Figure 6. Tree structures representing the position and posture of an object and their probabilities.

and branch nodes $b1, b2$. The parameters which connect with $a2, a3$ are restricted, and they are described by the tree structure. This processing is repeated until all sticks are described by the tree structure. At this point parameter tree which maintains stick connectivity is complete, with the reliance stick as the root (see Fig. 6). This tree represents the parameters group which is restricted by only the connectivity of sticks at the stage at which the parameters of rt were determined.

Next, a tree is generated using stick $i+1$ as the reliance stick rt (that is, the root), and a third tree is generated using stick $i+2$ as the root. Finally, multiple trees are generated having each stick as the root.

When all trees are completed, the probability of the parameters is added to each successive node further from the root. Then, the furthest nodes from the root have the total probability as shown in Fig. 6. The parameters which have the maximum values in all top-down paths (like the route expressed by the thick line in Fig. 6) can be considered as the parameters which correspond most accurately to the target among those parameter groups which maintain the connectivity of all sticks. Therefore, the probability distribution is reformed by placing these parameters at the center of the probability distribution.

Applying this method, the probability distribution can be reformed while maintaining the connectivity of sticks. Moreover, if the tracking in a stick fails because of occlusions, the parameters which fail can be recovered by relying on the parameters of other sticks. There are two reasons for selecting all sticks as the reliance stick. The first is that uncertainty remains if only one stick is selected as the reliance stick. The second is that the possibility of tracking failure becomes high when an occlusion occurs in the reliance stick.

3.4. Probability distribution propagation

The system propagates the probability distribution for a current image in order to obtain the probability distribution for a new image. The parameter setup of Φ is used to maintain the connectivity of sticks. The stick which has the maximum probability in all sticks is selected as the standard stick. The time when the current image is captured is expressed by t_n, and the time when new image is captured is expressed by t_{n+1}. The propagation of the probability distribution from t_n to t_{n+1} is computed by formula (10). Here, P^i_{peak} represents the peak of the probability distribution of stick i.

$$P^i_{peak}(t_{n+1}) = P^i_{peak}(t_n) + \frac{t_{n+1}-t_n}{t_n - t_{n-1}}\{P^i_{peak}(t_n) - P^i_{peak}(t_{n-1})\} \quad (10)$$

Since the variance of the distribution is changed during probability distribution reform, it is not changed during propagation. The scale of the distribution is also unchanged.

4. Experiments

To confirm the effectiveness of the proposed method, we selected goldfish as non-rigid objects lacking sufficient features, and conducted experiments with them.

4.1. Result of object model generation

Three hundred images (640×480pixel, YUV422, 16-bit color) captured for 10 seconds at 30 fps were used as the set of sample images for object model generation. We selected a set of images in which the goldfish were changing their postures, to learn the various postures.

In our study, the area of each goldfish was determined from red color information; in order to focus on pose estimation from silhouettes of observed objects, a background scene was simple and its color was very different from that of a target as shown in Fig. 7 and Fig. 9. After object model generation, the model with nine sticks shown in Fig. 4 was generated. The reliabilities of the sticks representing the fin of a goldfish were low. This is because self-occlusion occurred frequently at the fin. In contrast, the reliabilities of the ellipses representing the head were high, because the black portions of the eyes were rarer than the red portions.

4.2. Tracking experiments

We also conducted a tracking experiment using video (30fps, two goldfish) which was different from the sample images. Object models corresponding to

the peak of probability distribution were described in each frame during tracking. Only the stick models were described (i.e., without the ellipse model), in consideration of the ease of seeing the resulting images. The weights W_{clr}, W_{skl}, and W_{area} of each evaluation function in formula (8) were set to $W_{clr} = W_{area} = 2\,W_{skl}$, so that the weights of color histograms and areas would be high.

We conducted an experiment in which mutual occlusions occurred between targets. Fig. 7 shows an example of an acquired image sequence. In this experiment, the frame immediately before the occlusion occurred was used as the initial frame. Hypotheses A and B were generated in the initial frame #40, and the occlusion occurred in #52. Although hypotheses A and B corresponded to the postures of the objects in #62, they did not correspond accurately to the postures of the heads because of the occlusion in #67. In #77, hypothesis A was recovered for the head, by relying on the information from sticks which had not been affected by the occlusion. Hypothesis B was also recovered for the head, in the same way, in #82. Thus, even if the system cannot estimate the position and the posture of the targets accurately because of occlusions, it can estimate the targets accurately by recovering at a later point.

Transitions of the peak probability distribution of hypotheses A and B are expressed by the line graphs in Fig. 8. Frames in which the color is bright in the background represent estimations which failed. Judgment of estimation failure was made subjectively; the standards for judgment were whether all sticks were within the target areas, and whether all sticks corresponded to the posture of the targets.

Fig. 9 shows an example of estimation failures. Hypothesis A failed the estimation because of the occlusion and the change of posture in #66. Therefore, the probability was low around #66, as shown in Fig. 8 (top), but it was recovered by information regarding sticks which had not been affected by the occlusion in later frames. In #108, hypotheses A and B failed the estimation because of the occlusion, and in #136 hypothesis B failed the estimation because of the change of posture. Therefore, although the probabilities were low in these frames, we could recover the failure in later frames, as shown in 8 (top) and 8 (bottom). In #55, although hypothesis B failed the estimation because of the occlusion, the probability did not become low. It was considered that the head part of hypothesis B corresponded to the head of the goldfish of hypothesis A.

In this experiment, the rate of successful estimation was 87% in hypothesis A, 74% in hypothesis B, 81% on the average. Judgments of successful estimation were performed subjectively, as above. We judged a video of 10 seconds (300 frames), and occlusions occurred in 50 frames. Since the goldfish of hypothesis B was covered by that of hypothesis A, a difference occurred in the rate of successful estimation of hypotheses A and B.

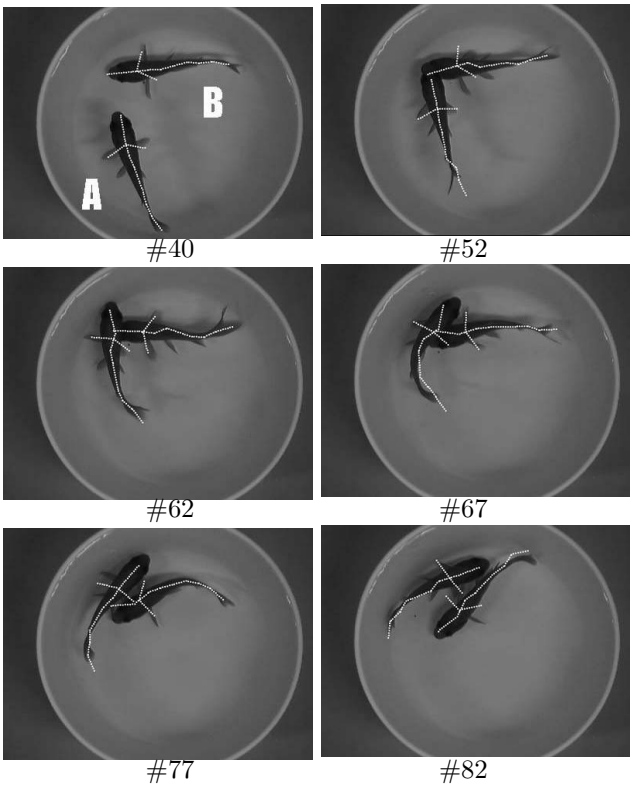

Figure 7. Experimental results.

We conducted a further experiment which the target size changed with the zoom function. In previous methods, since an object model has to be created subjectively, a great deal of labor is required. In our method, however, an object model corresponding to a change of target size could be generated automatically from sample images. The system tracked targets using an object model in which occlusions occurred. The same result was also obtained in this experiment, and we thus confirmed the effectiveness of the proposed method.

5. Conclusions and comments

We have proposed a non-rigid-object tracking method, with an object model generated automatically from a set of sample images. Moreover, we selected goldfish as non-rigid objects lacking sufficient features, conducted tracking experiments with them, and confirmed the effectiveness of our proposed method. Fu-

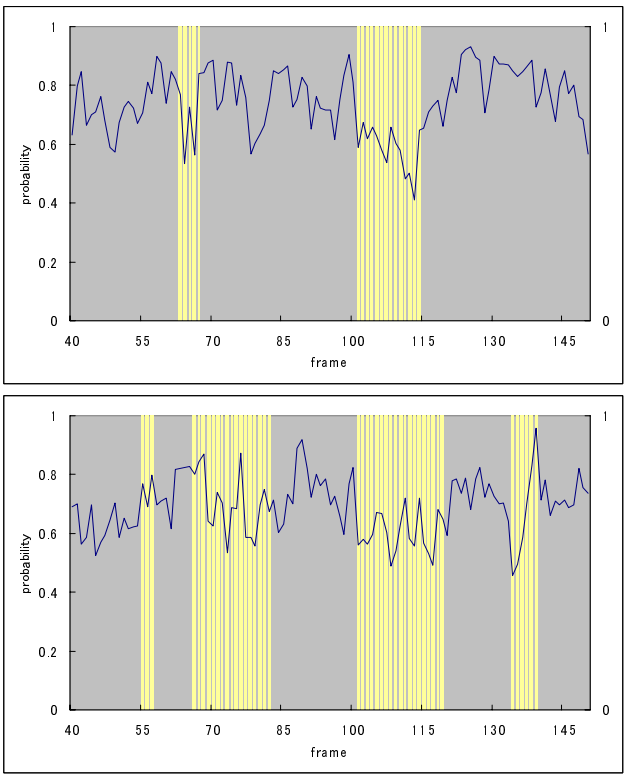

Figure 8. Transitions of the probability of hypotheses A (top) and B (bottom).

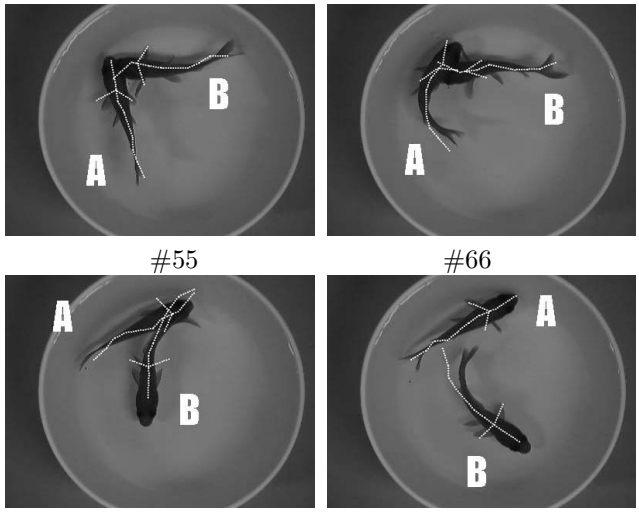

Figure 9. Examples of estimation failure.

ture work will include improvements in the estimation accuracy, processing time, and the flexibility of our method, by testing objects other than goldfish.

This work was supported by the PRESTO program of Japan Science and Technology Agency (JST).

References

[1] M. Isard and A. Blake, "Condensation - conditional density propagation for visual tracking", International Journal of Computer Vision, Vol.28, No.1, pp.5–28 (1998)

[2] J. MacCormick and A. Blake, "A probabilistic exclusion for tracking multiple objects", In *Proc. of International Conference on Computer Vision*, pp.572–578, 1999.

[3] A. Sugimoto, K. Yachi and T. Matsuyama, "Tracking Human Heads Based on Interaction between Hypotheses with Certainty", In *Proc. of The 13th Scandinavian Conference on Image Analysis*, pp.617–624, 2003.

[4] S. Kamijo, Y. Matsushita, K. Ikeuchi, and M. Sakauchi, "Occlusion Robust Tracking Utilizing Spatio-Temporal Markov Random Field Model", In *Proc. of International Conference on Pattern Recognition 2000*, Vol.1, pp142–147, 2000.

[5] S. Araki, N. Yokoya, and H. Takemura, "Tracking of multiple moving objects using split-and-merge contour models based on crossing detection", In *Proc. of Vision Interface*, pp.65–72, 1997.

[6] J. A. Sethian, "Level Set Methods and Fast Marching Methods", Cambridge University Press, 1999.

[7] I. Haritaoglu, D. Harwood, and L. S. Davis, "An appearance-based body model for multiple people tracking", In *Proc. of 15th International Conference on Pattern Recognition*, Vol.4, pp.184–187, 2000.

[8] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-rigid Objects using Mean Shift", In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition 2000*, pp.142–151, 2000.

[9] S. X. Ju, M. J. Black, and Y. Yacoob, "Cardboard people: A parameterized model of articulated motion", In *Proc of 2nd International Conference on Automatic Face and Gesture Recognition*, pp.38–44, 1996.

[10] Y. Kameda, M. Minoh, and K. Ikeda, "Three Dimensional Pose Estimation of an Articulated Object from its Silhouette Image", In *Proc. of Asian Conference on Computer Vision '93*, pp.612–615, 1993.