ROBUST OBJECT TRACKING USING A CONTEXT BASED ON THE RELATION OF OBJECT AND BACKGROUND

Takayoshi Yamashita

OMRON Corporation
2-2-1, Nishikusatsu, Kusatu-shi, SHIGA

Hironobu Fujiyoshi

Chubu University
1200, Matsumoto-cho, Kasugai-shi, AICHI

ABSTRACT

Object tracking is complicated by perspective changes to both the object and background caused by object and the camera motion, object-to-object and object-to-background occlusion and illumination changes. Conventional object tracking method focus on distinguishing the target from the background. Adopting a new perspective, we propose a context aware tracking by a collaborative model of both the object and its surrounding background. In this model we introduce the "probability of tracking failure" that determines the feature similarity and the spatial relationship of the target and the surrounding background, as the "target-surrounding context", which appears effective to predict the likelihood of a tracking failure. This target-surrounding context can be used to prevent tracking failure where the background has similar objects. In the experimental of the scene which occurs occlusion by similar objects, the proposed method outperformed most of the conventional methods.

***Index Terms*—** tracking, context, surrounding-region, vicinity, occlusion

1. INTRODUCTION

The object tracking is required to track continuously during appearance changes and occlusion happen under clutter background. Many of the existing methods have tried to track these situation, but occlusion still remain as one of the important problem [1][2][3][6]. To overcome this problem, Yang proposed a method that models the target and the associated stable background regions [5]. This method estimates the target object region from the temporal relationship with the persistent background region. Grabner proposed a method that estimates the target region from its temporal relationship with the invariant features of the background regions [7]. Fan propose a method that selects local regions in target as attention region which is discriminated from the background[8]. Dinh also proposed a method that focuses on background region which is similar with target object by detection approach[9]. These methods can even track occluded objects. But tracking fails for objects with similarities to the background or when occluded by similar objects as it is difficult to discriminate the target from the background. In this paper, we propose a method that is not target object centered but focuses on the effective use of information surrounding the target object. As a human being's, human attention increases to avoid lost when similar objects is in the background. We focus on this human visual characteristics which is influenced by the context and propose a method that tracks even when occluded by similar objects or background. The proposed method tracks both the target and surrounding regions to understand the target-surrounding context which is determined from the relationship between them.

2. RELATED OBJECT TRACKING MODEL

Yang proposed a model that employs background regions that appear steadily over time as auxiliary regions and considers the relationship between the target object and these auxiliary regions [5]. The posterior probability can be given by Eq.(1).

$$p(x_{t,0}|Z_{t,0}) \propto p_0(z_{t,0}|x_{t,0})p_0(x_{t,0}|Z_{t-1,0}) \prod_k m_{k0}(x_{t,0}) \quad (1)$$

Note that $x_{t,0}$ denotes the target position at time t and k denotes an auxiliary region ($k = 1, ..., K$). Additionally, $m_{k0}(x_0)$ indicates information passed to the target from auxiliary region k as given by Eq.(2).

$$m_{k0} = \int_{x_{t,k}} p_k(x_{t,k}|Z_{t,k})\psi_{k0}(x_{t,k}, x_{t,0})dx_k \quad (2)$$

$\psi_{k0}(x_{t,k}, x_{t,0})$ is the probability of a positional relationship between a continually appearing auxiliary region obtained by data mining and the target object. A region that is moving in the same way as the target object is often selected as an auxiliary region.

3. PROPOSED OBJECT TRACKING MODEL

3.1. Context aware tracking model

Yang infer the position of the target from the positional relationship between the target and auxiliary regions with continuously appearing, stable background regions. This method, however, goes no further than reflecting this positional relationship in tracking results and makes no attempt at understanding the immediate context. Drift can, therefore, occur if an object similar to the target is close to or occludes the target. To address the above problem, we propose an object tracking model that understands the context based on relationship between the target and the background. When a human being is focusing on a particular object, the degree of attention increases if similar objects are found in the vicinity. On the other hand, the degree of attention is lower the target object is distinct. To make use of this human visual characteristics for object tracking, it is important to understand the context surrounding the target. Inspired by this human visual characteristics, we define a adaptive tracking model based on the surrounding context.

$$p(x_{t,0}|Z_{t,0}) \propto p_0(z_{t,0}|x_{t,0}, c_{t,0})p_0(x_{t,0}|Z_{t-1,0}, c_{t,0}) \quad (3)$$

Here, $p_0(z_{t,0}|x_{t,0}, c_{t,0})$ is the observation model of the target at time t and $p_0(x_{t,0}|Z_{t-1,0}, c_{t,0})$ is the state model with respect to target x_t at time t. To obtain the probability $p(x_{t,0}|Z_{t,0})$, the observation and the state model use $c_{t,0}$ as a condition expressing the target-surrounding context of the target at time t. The target-surrounding context $c_{t,0}$ of the target is defined from the context of all surrounding regions as follows.

$$\begin{aligned} c_{t,0}(x_{t,0}) &= \prod_k p_k(x_{t,k}|Z_{t,k})(1 - \psi'_{k0}(x_{t,k}, x_{t,0})) \\ &- \prod_k p_k(x_{t,k}|Z_{t,0})\psi'_{k0}(x_{t,k}, x_{t,0}) \end{aligned} \quad (4)$$

Two models are applied to one surrounding region. The first model expressed by the first term of the equation(4) corresponds to tracking using the feature of the surrounding region, and the second model expressed by the second term corresponds to tracking based on feature of the target. That is, the first term tracks the surrounding region while the second term determines whether a region similar to the target exists in the vicinity of the target. Function $\psi'_{k0}(x_{t,k}, x_{t,0})$ is calculated based on the spatial relationship and feature similarity between the target and the surrounding region. If a similar object does not exist in the vicinity of the target, $\psi'_{k0}(x_{t,k}, x_{t,0})$ is small and $c_{t,0}$ is positive. Conversely, if a similar object exists in the vicinity of the target object, $\psi'_{k0}(x_{t,k}, x_{t,0})$ is large and $c_{t,0}$ is negative. As explained above it is possible to understand the "target-surrounding context" from the probability of tracking failure function $\psi'_{k0}(x_{t,k}, x_{t,0})$ with respect to surrounding region k.

3.2. Probability of tracking failure

As shown in Fig. 1, there are three important elements in understanding the target-surrounding context: the positional relationship S_d between the target object and the surrounding region (Fig.1(a)), difference in direction of motion S_m

(a)Distance (b)motion (c)feature similarity

Fig. 1. Relationship between target object and surrounding regions

(Fig.1(b)), and feature similarity S_a (Fig.1(c)). The probability of tracking failure with respect to the target can be described on the basis of these elements as shown by Eq.(5).

$$\psi'_{k0}(x_{t,k}, x_{t,0}) = 1 - \exp(-S_d\, S_m\, S_a) \quad (5)$$

Positional relationship S_d, direction of motion S_m, and feature similarity S_a are described by Eqs.(6), (7), and (8), respectively.

$$S_d = \begin{cases} 1 & ||x_{t,k} - x_{t,0}|| < T_d \\ 0 & otherwise \end{cases} \quad (6)$$

$$S_m = \begin{cases} 1 & \frac{x_{t,k}-x_{t-1,k}}{||x_{t,k}-x_{t-1,k}||} \cdot \frac{x_{t,0}-x_{t-1,0}}{||x_{t,0}-x_{t-1,0}||} > 0 \\ 0 & otherwise \end{cases} \quad (7)$$

$$S_a = \textstyle\sum_{b=1}^{B} Q(b) \ln \frac{Q(b)}{P(b)} \quad (8)$$

In Eq.(8), Q and P denote the feature of the surrounding region and the target, B is the number of elements of feature vector, respectively. Similarity S_a is computed on the basis of Kullback-Leibler (KL) divergence. Function $\psi'_{k0}(x_{t,k}, x_{t,0})$ is large, if similarity is high while the distance between the target object and surrounding region is less than a certain value and motion is in an approaching direction. In this case, $c_{t,0}$ takes on a negative value. For all other situations, $c_{t,0}$ takes on a positive value and it is understood that as context that the probability of tracking failure is low.

4. IMPLEMENTATION

4.1. Tracking Framework

To realize the context aware tracking model, we describe the novel framework considering relationship between target and surrounding regions as shown in Fig. 2. The observation model of target and surrounding regions are initialized at time $t = 0$. After time $t = 1$, target and surrounding regions are tracked based on their respective models. For each surrounding region, the models for both the target and the surrounding region are applied independently. Then, the relationship is calculated from both tracking results and the probability of tracking failure using Eqs. (5), (6), (7), and (8), and the target-surrounding context is derived from the all the relationship as in Eq.(5). Target position is estimated according to the target-surrounding context and observation models and state models are updated based on the context.

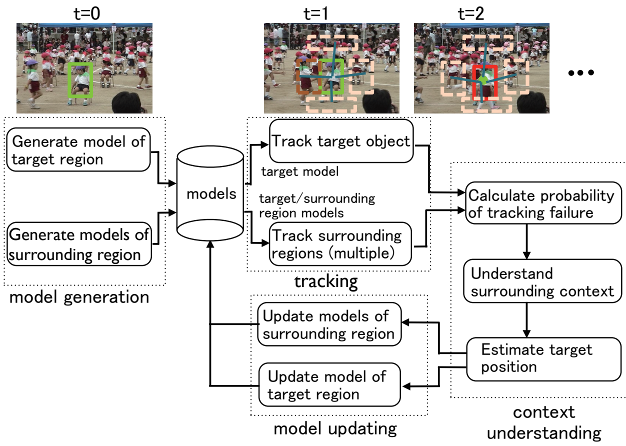

Fig. 2. Framework of proposed method

4.2. Observation model

We apply a joint histogram combining two co-occurrence histograms with hue-saturation and saturation-brightness in order to track nonrigid objects and under varying illumination. we employ kernel-based tracking for the similarity measurement between the target and candidate regions by KL divergence[6]. This has better tracking performance than mean shift based tracking.

4.3. Determination of target position from the context

Target position $x_{t,0}$ at time t is determined from the target-surrounding context. If $c_{t,0}$ in Eq. (4) is a positive value, it means a state in which the probability of tracking failure is low. Candidate position x_0^* is used as the target position $x_{t,0}$, $x_{t,0} = x_0^*$. Conversely, if $c_{t,0}$ is a negative value, it indicates a state where similar object is in the vicinity and the probability of tracking failure is high. Target position is determined from the surrounding regions as shown in Eq. (9).

$$x_{t,0} = \frac{1}{|D|} \sum_{x_k \in D} p(x_k|z_k) x_{k0}^* \tag{9}$$

Note that $|D|$ is the number of regions in subset D and x_{k0}^* is the target position estimated from surrounding region k. For subset D whose similarity to the target is less than the threshold T_a, $D = \{x_k | S_a(x_{t,k}, x_{t,0}) < T_a\}$.

4.4. Arrangement and Updating of surrounding regions

Surrounding regions are arranged radially at equal intervals from the target position. The regions are located above, below, left, and right of the target object in the case of 4 surrounding regions is used.

For effective selection of the surrounding region, the surrounding regions are generated and eliminated based on probability. For surrounding region x_k, tracking continues until probability $p(x_k|z_k)$ is lower than T_{pk}. If probability $p(x_k|z_k)$ drops and less than T_{pk}, new surrounding region will be generated. Additionally, if the distance between a surrounding region and the target is larger than twice of T_d, it will also be eliminated and a new surrounding region will be generated. This corresponds to the case in which no similar objects exist in the vicinity of the target object or in which the target or background is moving. On the other hand, if probability $p(x_k|z_0)$ is higher than T_{p0}, this surrounding region will be tracking continuously based on the feature of the target. Conversely, if probability based on the target is lower than T_{p0}, this surrounding region is eliminated and a new one generated. As shown above, surrounding regions are generated and eliminated in accordance with the context.

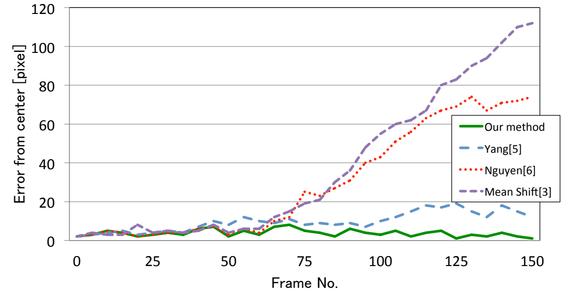

Fig. 3. Tracking error in the scene of target occlusion.

5. EVALUATION EXPERIMENTS

we compared the tracking performance with three conventional methods[3][5][6]. For the evaluation, we evaluated three home videos with similar objects overlapping, target object movement, and camera shaking. In our method, the joint histogram employed as an observation model consists of 16 $\times$ 16 bins. The number of surrounding regions is set to 4, and the parameters are as follows: given the size of the target object as S_0, T_d is 2.0 times S_0, T_{pk} and T_{p0} are 0.8, 0.7, respectively. The initial position of the target object is the same for all methods.

5.1. Performance in a scene with occlusion by a similar object

We compared performance in a 150-frame scene in which a person dressed in colors similar to the target occludes the target by passing in front of it. Tracking error from center position in each frame for the proposed method and other ones are shown in Fig. 3 and tracking results are shown in Fig. 4. In Fig.4(c), the red oval signifies a state in which the probability of tracking failure is high. In other words, a person dressed in color similar to the target object passes in front of it around the 70th frame. From this frame on, the methods not using the surrounding regions erroneously tracks the person passing in

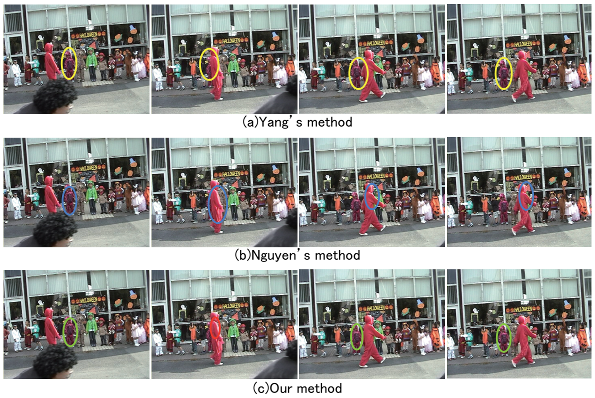

Fig. 4. Example of tracking result in the scene of target occlusion. First column is 60 frame, second column is 70 frame, third column is 80 frame and fourth column is 100 frame.

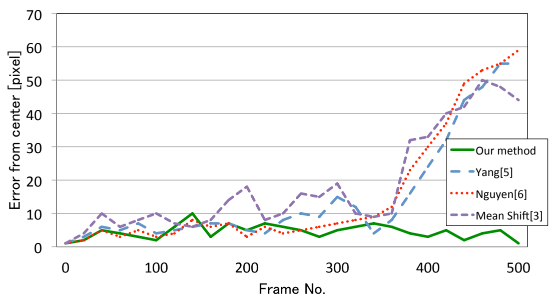

Fig. 5. Tracking error in the scene of crossing similar objects continuously under complex background.

front of the target object and as a result the error propagate. In contrast, the proposed method and in Yang's method, by virtue of estimating the position of the target object from the surrounding regions, correctly tracks the target object even after the similarly dressed person has passed the target object around the 80th frame.

5.2. Performance in a scene with continuous crossing of similar objects against a complex background

The performance in a scene with continuous crossing of similar objects against a complex background accompanied with movement and occlusion of the target object are shown in Fig. 5 and Fig. 6. The evaluation video consist of 600 frames and persons dressed similarly to the target object at a sports-day event are present in the background. In addition, the target object passes in front of a similarly dressed person and sits down on a chair, after which a similarly dressed person passes in front of the target object. Near the 200th frame, the target object passes in front of a similarly dressed person, and both the proposed technique and existing techniques continue tracking without switching from one object to another. Near the 350th frame, however, a similarly dressed person passes in front of the target object and both the method without surrounding regions and the Yang' s method drift to the passing person. The proposed method, meanwhile, can correctly track the target object even in such a complex scene. Yang's method cannot determine whether a similar object is present near the target object and cannot, as a result, determine whether it is tracking the target object or similar object in the event of occlusion. Moreover, objects in the background are moving resulting in no stable auxiliary regions as were present in the scene of subsection 5.1, which also causes drift in Yang's method. The proposed method derives the probability of tracking failure from the context, and if the probability of tracking failure happens to be high, it estimates the position of the target object based on those surrounding regions that are not similar to the target object. As a result, the position of the target object can be correctly tracked even under occlusion.

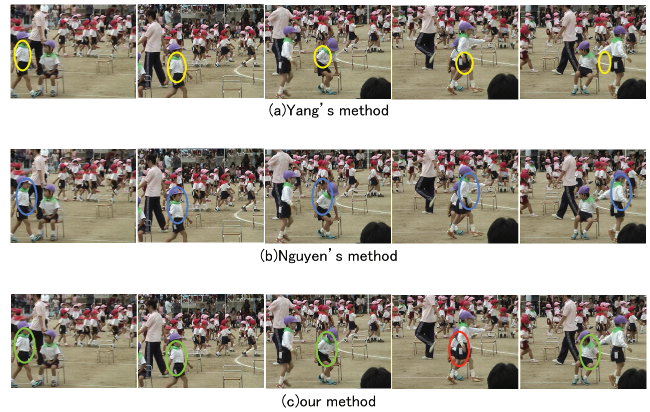

Fig. 6. Example of tracking result in crossing similar objects continuously under complex background . First column is 200 frame, second column is 300 frame, third column is 350 frame, forth column is 360 frame and fifth column is 370 frame.

6. CONCLUSION

We proposed a novel tracking model that focuses not just on the target object but also on target-surrounding context defined by surrounding regions as a new object-tracking technique. The proposed method seeks to understand the context in regions surrounding the target object and to change tracking behavior accordingly so that similar objects or background regions are not erroneously tracked. In an evaluation experiment, we compared this method with other methods focusing on scenes in which the target object coexists with similar objects and background regions. The results of this experiment showed that the proposed method can tracking a object even in complex scenes with similar objects occluding or in the near vicinity.

7. REFERENCES

[1] M. J. Black and A. D. Jepson. Eigentracking: Robust matching and tracking of articulated objects using a view-based representation. International Journal of Computer Vision, Vol.26, No.1, pp.63–84, 1998.

[2] M. Isard and A. Blake. CONDENSATION—Conditional Density Propagation for Visual Tracking. International Journal of Computer Vision, Vol.29, No.1, pp.5–28, 1998.

[3] D. Comaniciu, V. Ramesh, P. Meer. Kernel-based object tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.25, No.5, pp.564–577, 2003.

[4] C. Yang, R. Duraiswami, D.Larry. Efficient mean-shift tracking via a new similarity measure. IEEE Conference on Computer Vision and Pattern Recognition, pp.176–183, 2005.

[5] M. Yang, Y. Wu, S. Lao. Intelligent Collaborative Tracking by Mining Auxiliary Objects. IEEE Conference on Computer Vision and Pattern Recognition, pp.697–704, 2006.

[6] Q. Nguyen, A. Robles-Kelly, C. Shen. Kernel-based Tracking from a Probabilistic Viewpoint. IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2007.

[7] H. Grabner, J. Matas, L. Van Gool. Tracking the invisible: Learning where the object might be. IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2010.

[8] J. Fan, Y. Wu, S. Dai, "Discriminative spatial attention for robust tracking", European Conference on Computer Vision, 2010.

[9] T. B. Dinh, N. Vo, G. Medioni, "Context Tracker: Exploring Supporters and Distracters in Unconstrained Environments" IEEE Conference on Computer Vision and Pattern Recognition, pp.1–8, 2011.