Tracking Failure Detection by Imitating Human Visual Perception

Hyung Jin Chang¹, Myoung Soo Park², Hawook Jeong¹, Jin Young Choi¹

¹Perception and Intelligence Laboratory
School of Electrical Engineering and Computer Science, ASRI
Seoul National University, Seoul, Korea
²Korea Institute of Science and Technology, Seoul, Korea

Abstract

In this paper, we present a tracking failure detection method by imitating human visual system. By adopting log-polar transformation, we could simulate properties of retina image, such as rotation and scaling invariance and foveal predominance. The rotation and scaling invariance helps to reduce false alarms caused by pose changes and intensify translational changes. Foveal predominant property helps to detect the tracking failing moment by amplifying the resolution around focus (tracking box center) and blurring the peripheries. Each ganglion cell corresponds to a pixel of log-polar image, and its adaptation is modeled as Gaussian mixture model. Its validity is shown through various experiments.

Index Terms— Tracking failure detection, Human visual system, Log-Polar transform, Gaussian Mixture Model

1. Introduction

In computer vision, there have been lots of efforts to improve tracking performance, and as a result, most algorithms work well for many challenging situations. Nevertheless, they still lost their tracking object in the long run. Current visual surveillance system [1] restores failed tracker manually. However if we can detect a tracking failure moment, the restoration can be performed automatically. So the tracking failure detection (TFD) is an important component for automatic tracking system.

Most of the existing TFD methods are based on checking similarity measures. [2, 3] detect tracking failure by thresholding a similarity measure of tracker. However, because the similarity measures are not originally designed for TFD, they cannot represent a status of current tracker exactly. Sometimes the similarity measure frequently results in a low value even when the tracking is successful or varies smoothly when the tracking fails by slow changes. So [4] defines a new similarity measure only for TFD. It is assumed that the boundary of tracking box does not include any pixels of tracking object. However, in actual application, this assumption may be easily violated and as a result it leads to frequent false alarms.

We propose a new approach for TFD by mimicking human visual system. When people look at an object, the attentional area seems clear but peripheries are blurred. This is because of the structure of the retina of a human eye. Fovea [5] is a part of the eye, located in the center of the macula region of the retina. The fovea is responsible for sharp central vision and is surrounded by the parafovea belt and the perifovea outer region. The parafovea and perifovea are composed of sparse ganglion cells [5]. Approximately 50% of the nerve fibers in the optic nerve carry information from the fovea, while the other 50% carry information from the rest of the retina. Log-polar image geometry was first motivated by its resemblance with the structure of the retina [6]. We use a log-polar image for simulating human vision and its characteristics.

Our method focuses on capturing a distinctive feature when tracking fails instead of comparing similarity measures. At the instant when the target object moves out of the area of focus (i.e. tracking fails), the object suddenly becomes blurry, whereas the surroundings gets sharp. We use this sudden sharp and blur view change as an important feature for TFD. Human can detect the changing moment by perceiving the amount of fired (stimulated by new color) ganglion cell. We model ganglion cells in the retina as pixels in log-polar image and the adaptation of ganglion cell as Gaussian mixture model (GMM) [7]. So, the perception of the amount of fired ganglion cells is modeled by counting new colored pixels in the log-polar image. This measure is independent of the tracker, so it can be applied to any trackers. The effectiveness of the proposed TFD is shown by several experiments.

2. Characteristics of Log-Polar Image and Tracking Failure

2.1. Properties of Log-Polar Image

The log-polar transformation [6] means a conformal mapping (preserves oriented angles between curves and neighborhood relationships) from the point \((x, y)\) on the cartesian plane to point \((\rho, \theta)\) in the log-polar plane, where
Definition 2.1 (Tracking failure) The tracking failure moment is defined as the moment when the center of tracking object ($C_T$) is moved to background region ($R_B$) from the region of tracking object ($R_T$).

The $C_T$ corresponds to foveated point of eye. According to Definition 2.1, tracking failure appears when the $C_T$ crosses over the boundary line between $R_T$ and $R_B$. It means that, under our definition, translational changes are more important than rotational and scaling changes in the tracking box images.

In Fig. 3, we model the tracking failure situation. The result shows that log-polar transformed image intensifies the changes around $C_T$ and decrease the changes of peripheries. This is induced by nonlinear predominance property of log-polar transformation and it helps to capture a boundary crossing moment and ignore other background changes. So, the two properties (rotation and scaling invariance, and foveal predominance) of log-polar image are effective for TFD.

3. TRACKING FAILURE DETECTION ALGORITHM

3.1. Modeling of Ganglion Cell Adaptation

From the biological plausibility of log-polar image, each ganglion cell corresponds to each pixel of log-polar image. For dynamic modeling of pixels in tracking box image, we adopt the framework of online GMM method [7].

$$P(X_n(t)) = \sum_{k=1}^{K} \omega_n^k(t) \ast \eta(X_n(t), \mu_n^k(t), \Sigma_n^k(t))$$

where $\omega_n^k(t)$ is an weight, $\mu_n^k(t)$ is the mean value and $\Sigma_n^k(t)$ is the covariance matrix of each Gaussian in the mixture at time $t$ respectively. In this formulation, $\omega_n^k(t), \mu_n^k(t), \sigma_n^k(t)$ are updated by following equations as in [7].

$$\omega_n^k(t) = (1 - \alpha)\omega_n^k(t - 1) + \alpha M_n^k(t)$$

$$\mu_n^k(t) = \frac{1}{P_n(t)} \sum_{X_n(t) \in R_n} \omega_n^k(t) X_n(t)$$

$$\Sigma_n^k(t) = \frac{1}{P_n(t)} \sum_{X_n(t) \in R_n} \omega_n^k(t) (X_n(t) - \mu_n^k(t))(X_n(t) - \mu_n^k(t))^T$$

$$P_n(t) = \sum_{k=1}^{K} \omega_n^k(t)$$

$$M_n^k(t) = \frac{1}{P_n(t)} \sum_{X_n(t) \in R_n} \omega_n^k(t) X_n(t)^T$$

where $\omega_n^k(t)$ is an weight, $\mu_n^k(t)$ is the mean value and $\Sigma_n^k(t)$ is the covariance matrix of each Gaussian in the mixture at time $t$ respectively. In this formulation, $\omega_n^k(t), \mu_n^k(t), \sigma_n^k(t)$ are updated by following equations as in [7].
where $\alpha$ is a learning rate and $M_h^k(t)$ is 1 for a matched model and 0 for the others.

$$
\mu_k^h(t) = (1 - \nu)\mu_k^h(t - 1) + \nu X_n(t)
$$

$$
\sigma_n^k(t)^2 = (1 - \nu)\sigma_n^k(t - 1)^2 + \nu (X_n(t) - \mu_n^k(t))^T (X_n(t) - \mu_n^k(t)),
$$

where $\nu = \alpha\eta(X_n(t)\mu_n^k(t), \sigma_n^k(t))$.

We set initial values of $\omega_{init}(1), \mu_{init}(1), \sigma_{init}(1)^2$ using the color information of initial tracking box image $X(1)$. Because we do not know how many colors the tracking object are composed of, mean shift clustering (MSC) [8] method is used to find the number.

With $N$ pixel points \{ $X_1(1), \ldots, X_N(1)$ \} $\in R^3$ (RGB color space), we find $K$ clusters ($K$ color distributions) by means of MSC. Each color distribution $C_k(k=1\ldots K)$ is composed of $n_k \sum_{k=1}^{K} n_k = N$ pixel points $X_i^{k}(1)$. We model each color distribution as Gaussian distribution and calculate initial parameter values with clustered pixel points. $\omega_{init}(1) = n_k/N$ is an weight, $\mu_{init}(1) = (\sum_{i=1}^{n_k} X_i^{k}(1))/n_k$ is the mean. $\Sigma_{init}(1) = \sigma_{init}(1)^2 I$ (each color space is independent and have the same variance $\sigma_{init}(k)^2 = (\sum_{i=1}^{n_k} (X_i^k(1) - \mu_{init}(1))^2)/n_k$ is the covariance matrix of each $k^{th}$ ($k = 1 \ldots K$) color distribution respectively. $N$ pixels of $X(t)$ share the same initial values.

### 3.2. Tracking Failure Detection

Every ganglion cell is fired (stimulated by new color) independently. New color perception is modeled as checking abruptly changing pixel (ACP).

$$
ACP_n(t)=\begin{cases} 
0 & \text{if} \ (X_n(t) - \mu_n^k(t))^2 < 2.5\sigma_n^k(t)^2, \\
1 & \text{otherwise}.
\end{cases}
$$

Then, in order to detect tracking failure, ACP ratio $\xi_{ACP}$ in current image $X(t)$ is measured by

$$
\xi_{ACP} = \frac{\sum_{n=1}^{N} ACP_n}{N}.
$$

Using $\xi_{ACP}$, tracking failure is determined by thresholding:

$$
\chi_{TFD}(X(t)) = \begin{cases} 
1 & \text{if} \ \xi_{ACP} > T; \\
0 & \text{otherwise}.
\end{cases}
$$

where $T$ is a threshold value, experimentally defined.

### 4. EXPERIMENTAL RESULTS

To evaluate the validity of our TFD algorithm, we conducted some experiments. We implemented our algorithm in MATLAB for simulation with a threshold $T = 0.4$.

![Fig. 4](image4.png) (a) shows the $\xi_{ACP}$ comparison. Occlusion occurs as frame 56. (b) is the tracking object image of frame 1. (c) and (d) are images of frame 58 and its ACP image in Cartesian space and log-polar space respectively.

![Fig. 5](image5.png) Fig. 5. The first row represents a TFD result without initial color model generation and the second row is a TFD result by using initial color model.

### 4.1. Effectiveness of Log-Polar Transformation and Initial Color Model Generation

We verify our claim that log-polar space is suitable for TFD than cartesian space. Fig. 4 shows a comparison between the ACP detection in two different spaces, cartesian space and log-polar space. As we can see in Fig. 4(a), the change around $C_{TB}$ is magnified in log-polar space.

Fig. 5 shows the effect of setting initial color model. There are several inner boundaries in $R_{TFD}$ which induce false alarms. By setting initial color model for GMM, we could achieve to give less alarms for inner boundaries.

To evaluate the performance of the proposed algorithm, we compare the TFD accuracy with K-means Tracker TFD [4] (Because [4] is a tracker independent TFD measure based method same as ours). As we can see in Fig. 6, our method can afford to occlusion and scale changes not giving false alarm until tracker really misses the target.

### 4.2. Combining with various tracking algorithms

The proposed TFD method can be applied to any tracking algorithm. Fig. 7 shows combined TFD results with different tracking methods, kernel based tracking [9] and particle filter tracking [10]. Because our method evaluates current tracking status not by an implicit similarity measure of tracker but by an explicit tracking result image (which is analogous to the way of people make a decision), we can see that our TFD method can be successfully combined with any kinds of tracking algorithm.
Fig. 6. K-means TFD gives an alarm when the score is over 0.7 (This value is from [4]), and the proposed TFD gives alarm when the score is over 0.4. In ground truth, tracking fail occurred at frames 35 in (a) because of occlusion and frames 42 in (b) because of target becomes too small.

Fig. 7. The first row shows TFD results combined with kernel based tracking [9] and the second row is the TFD result with particle filter based tracking [10].

The proposed method also can be used for enhancing tracking performance by feedback. Fig. 8 shows tracking performances of IVT tracker [10] measured by root mean square error (RMSE) comparing to ground truth. When tracking failure measure increases, TFD makes the tracker to stop updating tracking template models and widen particle spreading range. This feedback helps tracking algorithm more robust.

5. CONCLUSION

In this paper, we proposed a tracking failure detection method by mimicking human visual system. By adopting log-polar transformation for modeling retina image, we could intensify the translational change. This property makes it possible to detect tracking failure moment easily. We modeled ganglion cell adaptation using online GMM and detected abrupt change in pixels. Experimental results shows that our method gives less false alarms, and can be applied to any tracking methods.

6. ACKNOWLEDGMENT

The research was sponsored by Samsung Techwin Co.,Ltd. and SNU Brain Korea 21 Information Technology program.

7. REFERENCES


