NOVEL COLOR-BASED TARGET REPRESENTATION FOR VISUAL OBJECT TRACKING

Yuantao Gu *, *Yilun Chen*

Department of Electronic Engineering
Tsinghua University
Beijing, 100084, CHINA
Email: gyt@tsinghua.edu.cn

Jie Wang

Dept. of Electrical and Computer Engineering
University of Toronto
Toronto,ON, M5S3G4, CANADA
Email: jwang@dsp.utoronto.ca

ABSTRACT

In this paper, a novel color-based target representation scheme is proposed for object tracking. Different from those commonly used models, where color-based features are extracted from the object region only, the proposed solution takes the background information into account as well. A transition region is defined which contains the background area around the target object. Thus the target object is represented by the color distribution estimated from both the object region and the transition region. Negative weights are assigned to the pixels in the transition region so that only the distinct features which are distinguishable from the background are extracted to represent the target object. Experimental results suggest that the proposed model outperforms the traditional model in the scenarios where surrounding background is similar to the target object or each part of the object has similar appearance.

1. INTRODUCTION

Visual object tracking is one of the most fundamental tasks in the applications such as multimedia surveillance, human computer interaction, teleconferencing, video editing and compression [1]. The objective of visual tracking is to iteratively find the given object in an image sequence. In general, a complete object tracking system is formed by two key modules, target representation and data association [2]. The objective of target representation is to extract object features which describe the appearance of the specific target. The task of data association is to analyze the dynamic status of the target object and predict the corresponding position. Many data association techniques have been proposed in literature, with the state-of-the-art such as Mean Shift [3] and Particle Filter [4, 5, 6, 7]. In this paper, we focus on the target representation module and the Particle Filter is selected as the data association technique.

The target object can be represented by using various features, such as color [2, 4, 8], contour [5, 6, 7], and feature points [9]. In general, color is one of the most attractive features which is commonly used in literature due to its robustness against non-rigidity, rotation, and partial occlusion. In the current color model, the target object is usually represented by the color distribution estimated from the object region which is usually bounded by an ellipse. In addition, a kernel function is applied to assign smaller weights to the pixels farther from the center of the object region. Such color distribution based model is widely used in literature to represent the target objects such as face and car [2, 4].

*This project is partially supported by National Science Foundation of China (NSFC 60402030), and Development Research Foundation of the Department of Electronic Engineering, Tsinghua University.

Although many successful examples have been reported in literature by using the above color model [2, 4], it usually fails in the following scenarios which are often encountered in practice: (1) the color distribution of background is similar to that of the object. In such a case, background region may be mis-included, thereafter, the estimated object region deviates from its real position. The continuous deviation may lead to an object missing; (2) some parts of the object have similar color distribution to that of the whole object, especially for compact objects such as face. In such a case, the tracking may be stuck into those similar sub-regions. The tracking failure under these two scenarios is due to the fact that only object information is considered in the current model.

In order to solve the above mentioned problems, a novel color-based target representation solution is proposed in this paper. In the proposed color model, color distribution is estimated from both the pixels in object region and surrounding background which is denoted as transition region. A new kernel function is proposed which assigns negative weights to the pixels in the transition region. In such a case, the color components which may appear in both the object region and the background have less contribution to the color distribution estimation. Therefore, only the color information distinguishable from the background is extracted for target representation. Experimental results indicate that the proposed color model helps to eliminate similar background influence as well as to avoid the confusion of the object's sub-regions.

The rest of this paper is organized as follows. Traditional color-based target model is briefly reviewed in section 2 and the proposed target model is introduced and analyzed in section 3. In section 4, two sets of experiments are performed with the results presented in details. Conclusions are summarized in section 5.

2. COLOR-BASED TRACKING SYSTEM

In the current color-based tracking system, color distribution is widely used to represent the target object [2, 4, 8]. In general, the color distribution in a specific color space is expressed as a normalized histogram, i.e., $\mathbf{q} = [q_1, q_2, \cdots, q_M]^{\mathrm{T}}$, $\sum_{m=1}^{M} q_m = 1$ where M denotes the number of color bins and q_i denotes the normalized number of pixels whose color belongs to ith bin. Therefore, the reference target model is represented by the color distribution in the object region J [2, 4]:

$$Q_m = \sum_{\mathbf{s}\in J} k(\mathbf{s})\,\delta[h(\mathbf{s}) - m], \quad q_m = \frac{Q_m}{\sum_{m=1}^{M} Q_m} = \frac{Q_m}{\sum_{\mathbf{s}\in J} k(\mathbf{s})} \tag{1}$$

where $\delta(\cdot)$ is the Kronecker delta function and $h(\mathbf{s})$ is a function which assigns the color at location $\mathbf{s} = [x, y]^{\mathrm{T}}$ to a corresponding

bin. The kernel function $k(\mathbf{s})$ in Eq.(1) assigns smaller weights to the pixels farther from the center of J and higher weights to the pixels nearer to the center.

While in the tracking session, the target location in the current frame is firstly predicted from the previous frames by using a specific data association algorithm, such as Particle Filter. Then the color distribution of the candidates is calculated as follows:

$$p_m(I) = \frac{\sum_{\mathbf{s}\in I} k(\mathbf{s})\,\delta[h(\mathbf{s}) - m]}{\sum_{\mathbf{s}\in I} k(\mathbf{s})}, \tag{2}$$

where I denotes a hypothetical target region. The final determination of the target region is obtained by comparing the Bhattacharyya distance between the distributions of the candidates $\mathbf{p}(I)$ and that of the pre-defined reference target model $\mathbf{q}$, i.e.,

$$d(I) = \sqrt{1 - \sum_{m=1}^{M} \sqrt{p_m(I)\, q_m}}. \tag{3}$$

In the current object tracking algorithms, object region is usually bounded by an upright ellipse, i.e.,

$$I = \{\mathbf{s} | r(\mathbf{s}) \le 1\}, \tag{4}$$

where $r(\mathbf{s}) = \sqrt{\left(\frac{x-x_c}{a}\right)^2 + \left(\frac{y-y_c}{b}\right)^2}$ is the normalized distance and $[x_c, y_c]^T$, $[a, b]^T$ denote the center position and half axes of the ellipse respectively . The kernel function is usually defined as

$$k(\mathbf{s}) = \begin{cases} 1 - r^2(\mathbf{s}) & \text{if } r(\mathbf{s}) \le 1; \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Since color distribution is estimated from the object region only, it is incapable of discriminating the object from the similarly appeared background. In addition, if some parts of the object have the similar color distribution to that of the whole object (e.g., the color distribution of the cheek is similar to that of the whole face.), with the current color model, those similar sub-regions may be mis-captured instead of the object. In order to solve these problems, we proposed a new color model which takes the background information into account. More details will be discussed in the followed section.

3. NEW TARGET MODEL

3.1. Definition of the model

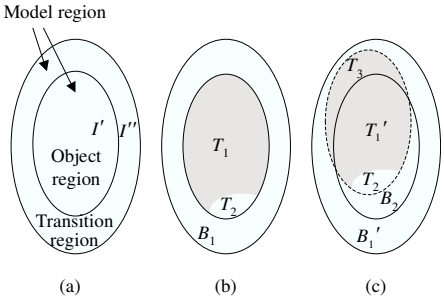


Fig. 1. (a) The proposed target representation model; (b) The reference of a general target ($T_1 + T_2$); (c) A candidate of (b).

Let's first discriminate two concepts: object region (I') and model region (I). Here, we denote the object region as the area including the target object while the model region is the area where the object features (e.g., color distribution) are extracted from. Obviously, in the traditional color-based tracking system, model region is identical to object region ($I = I'$), i.e., object features used for tracking are extracted from the object only. However, in the proposed here color model, color distribution is estimated not only from the object region but also its surrounding background, which is denoted as transition region (I''). Fig.1(a) depicts the proposed color model by using two ellipses. The smaller ellipse denotes the object region I' while the ring between two ellipses denotes the transition region I''. Therefore, the model region I is comprised of the object region I' and the transition region I'', i.e.,

$$I = I' \cup I'',\ I' = \{\mathbf{s} | r(\mathbf{s}) \le 1\},\ I'' = \left\{\mathbf{s} | 1 < r(\mathbf{s}) \le \sqrt{2}\right\}. \tag{6}$$

In order to discriminate the background and the object, a novel kernel function is proposed as follows,

$$k(\mathbf{s}) = \begin{cases} 1 - r^2(\mathbf{s}) & \text{if } r(\mathbf{s}) \le \sqrt{2}; \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where the boundary of $\sqrt{2}$ is determined to satisfy the condition $\int_{\mathbf{s}\in I} k(\mathbf{s}) = 0$.

Compared to the traditional kernel function where only object region is considered, the proposed function also takes the background region into account. It can be observed from Eq.(7) that the weights are negative if $1 < r(\mathbf{s}) \le \sqrt{2}$. Therefore, the reliability of the bins which the background colors belong to is reduced. This indicates that the colors appearing in both object region and background are not reliable to represent the object. Consequently, only discriminated features are extracted which helps to reduce the possibility of the mis-tracking to similar background regions.

In addition to the novel kernel function, another novelty of the proposed model exists in the definition of the color distribution which is obtained by estimating the color histogram from the model region (I) while normalizing it over the object region (I'), i.e.,

$$p_m = \begin{cases} p'_m & \text{if } p'_m > 0; \\ 0 & \text{otherwise,} \end{cases} \quad p'_m = \frac{\sum_{\mathbf{s}\in I} k(\mathbf{s})\,\delta[h(\mathbf{s}) - m]}{\sum_{\mathbf{s}\in I'} k(\mathbf{s})}. \tag{8}$$

Please note the denominators in Eq.(8) and Eq.(2), it can be easily observed that only the weights in I' are counted in Eq.(8), however, in Eq.(2), the normalization is performed over I. With such definition, the color distribution of the whole object and that of the object's sub-region with similar appearance can be discriminated. The detailed analysis is described in the followed subsection.

Since the normalization is not performed over the whole model region where the histogram is estimated, the proposed p_m is not a strictly defined distribution function due to the fact that $\sum_{m=1}^{M} p_m \le 1$. However, this will not affect the similarity computation of Eq.(3). For convenience purposes, it is still denoted as distribution.

3.2. Characteristics of the model

In this section, some analytical discussion will be provided to demonstrate how the proposed model solves the problems of similar background confusion and self-similarity.

A general tracking example is illustrated in Fig.1(b) and (c). Fig.1(b) denotes a reference target model, which is assumed to perfectly capture the target object. The target object is comprised of two parts, T_1 and T_2. We further assume T_1 has the distinct color distribution of the target, $\mathbf{t} = [t_1, t_2, \cdots, t_M]^{\mathrm{T}}$, however, the color distribution of T_2 is same as that of surrounding background, $\mathbf{b} =$

$[b_1, b_2, \cdots, b_M]^{\mathrm{T}}$. The two color distributions are further assumed to be irrelevant, i.e., $t_m b_m = 0$, $\forall 0 < m \leq M$. Please note that $\mathbf{t}$ and $\mathbf{b}$ are probability distribution of color bins, which means the calculated color distributions by Eq.(8) are also random variables, here their expected values are used for further analysis. According to the definition of Eq.(8) and using some basic statistics and geometry, the expected color distribution of the reference model is computed as follows:

$$\mathrm{E}\{q_m\} = t_m \frac{A_k(T_1)}{A_k(T_1 \cup T_2)}, \tag{9}$$

where $A_k(T) = \sum_{\mathbf{s}\in T} k(\mathbf{s})$ denotes the weighted area in a particular region. Please note that the color distribution of T_2 is the same to the background which indicates the color contribution from region T_2 is subtracted off by the transition region (refer to Eq.(8)), therefore, we got $A_k(T_1)$ on the numerator of Eq.(9) rather than $A_k(T_1 \cup T_2)$.

Fig.1(c) demonstrates a tracking situation in a specific frame. It can be observed that the candidate region (two solid ellipses) is not precisely coincident with that of the actual object region (dashed ellipse), i.e., T_3 is not included in the object region I' while the surrounding background B_2 is mis-included in I'. The color distribution representing the hypothetical candidate can be derived from the model region which is covered by the two solid ellipses as follows:

$$\mathrm{E}\{p'_m\} = \frac{t_m A_k(T'_1 \cup T_3) + b_m A_k(B'_1 \cup B_2 \cup T_2)}{A_k(T'_1 \cup T_2 \cup B_2)}. \tag{10}$$

Then the similarity between the candidate and the reference target can be computed. Please note in the second item of the numerator in Eq.(10), some of the color bins may have negative values due to the negative weights assigned to the pixels in I'', thereafter, a nonlinear mapping should be applied to make the negative values 0. However, in this example, such nonlinear mapping can be circumvented without affecting the final result, i.e., p'_m is used for similarity computation instead of p_m. This is due to the irrelevant assumption of $t_m b_m = 0$. So,

$$\mathrm{E}\{p'_m\}\,\mathrm{E}\{q_m\} = \frac{t_m^2 A_k(T'_1 \cup T_3) A_k(T_1)}{A_k(T'_1 \cup T_2 \cup B_2) A_k(T_1 \cup T_2)}. \tag{11}$$

Therefore, the expected Bhattacharyya distance can be approximated as

$$\begin{aligned} \mathrm{E}\{d\} &\propto \sqrt{1 - \sum_{m=1}^{M} \sqrt{\mathrm{E}\{p'_m\}\,\mathrm{E}\{q_m\}}} \\ &= \sqrt{1 - \sqrt{\frac{A_k(T'_1 \cup T_3) A_k(T_1)}{A_k(T'_1 \cup T_2 \cup B_2) A_k(T_1 \cup T_2)}}}. \end{aligned} \tag{12}$$

To simplify the derivation and make Eq.(12) more intuitively, we further suppose that the kernel function is normalized to

$$k(\mathbf{s}) = \begin{cases} 1 & \text{if } 0 < r(\mathbf{s}) \leq 1; \\ -1 & \text{if } 1 < r(\mathbf{s}) \leq \sqrt{2}; \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

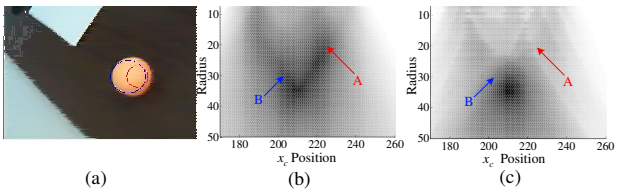
Then using the basic geometry, $T'_1 = T_1 - T_3$, $B'_1 = B_1 - T_3$, $B_2 = T_3$, we get

$$\mathrm{E}\{d\} \propto \sqrt{1 - \sqrt{\frac{(T_1 - 2T_3) T_1}{(T_1 + T_2)^2}}}. \tag{14}$$

Equation (14) explains why the proposed model works well in the scenarios that the traditional model fails: (1) T_2 denotes the part of the target subject but is similar to the background. As the area of T_2 increases, the distance from the candidate to the reference ($\mathrm{E}\{d\}$) will increase. Therefore, the tracker will select the candidate with small T_2, i.e., only the unique feature of the target is captured. (2) As T_3 increases, the corresponding T'_1 decreases, i.e., the tracker is stuck into a sub-region which has the similar color to the object. However, the increasing of T_3 leads to a larger $\mathrm{E}\{d\}$. Therefore, the mis-tracking of the object's sub-regions can be avoided. The following example will give the more obvious illustration of self-similarity avoidance.

3.3. An example

This example is to illustrate how the proposed model discriminates the whole object and the object's sub-region which has similar color distribution. Fig.2(a) is a static image containing a yellow table tennis ball centered at $[x_c, y_c]$ as the target. The ball is a compact object with a single color. Other than ellipse, spherical boundary is used in the color model. We manually select some candidates to compare their similarities to the reference target. The candidates are assumed to be centered at $[x, y_c]$ with the fixed ordinate y_c. However, x and the radius vary. Fig.2(b) and (c) depict the similarity values between various candidates and the reference target by using the traditional model and the proposed model respectively. The darker color denotes high similarity. In can be easily observed from Fig.2(c) that only the candidates in the actual object region have high similarity values. However, if the tradition model is used, some sub-regions of the object may be tracked with high similarity values. Please note, points A and B in Fig.2(b) which have the high similarity values are corresponding to the candidates with red circle and blue circle in Fig.2(a) respectively. This indicates, if the object is of self-similarity, the traditional model can not discriminate those similar sub-regions with the overall object. However, the proposed solution successfully eliminates those influences.



(a) (b) (c)

Fig. 2. (a) A static image with a distinct target whose color distribution is similar to that of its part region. (b) The similarity between various candidates and the reference model by using traditional model. (c) The similarity between various candidates and the reference model by using the proposed model.

4. SIMULATION RESULTS

In order to demonstrate the effectiveness of the proposed target representation model, a *face* sequence of 200 frames and a *car* sequence of 197 frames [10] are used for experimentation. The human face and the car are the corresponding target objects to be tracked. Both the proposed color model and the traditional color model [2, 4] are applied for comparison purposes. In addition, particle filter is utilized as the tracking algorithm for both sequences. The experiments are repeated 10 times and the reported here results are the average of 10 trials.

In the *face* sequence (Fig.3), a person is sitting and walking around in an uncontrolled room. The image size of each frame is

320x240 pixels. The initial face region is manually located, which is bounded by an ellipse with the axes of 37 and 55 pixels respectively. The initial face region is used to generate the reference target model. Fig.3 depicts the tracking results by using the traditional model (in red) and the proposed here model (in blue). It is well-known that skin color is evenly distributed all over the face, i.e., the color distribution of any part of the face is similar to that of the whole face. The advantage by introducing the proposed model can be easily observed. It can be observed that with the proposed model, the whole face region is exactly tracked. However, the traditional model tends to capture the sub-region of the face. Please note, in frame 72, the proposed model tracked the whole face region, whereas the traditional one only captures the sub-region around nose and mouth.

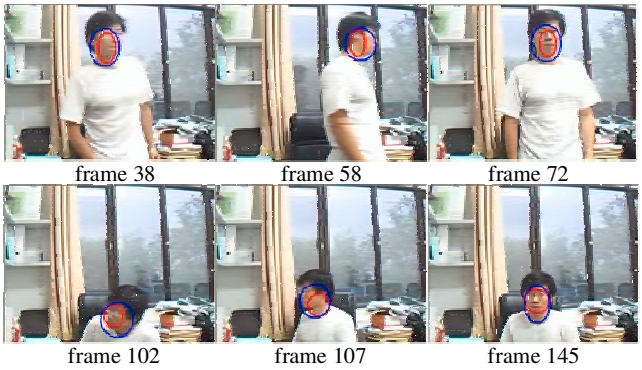


Fig. 3. Comparison of the face tracking performance of two models. (red: the traditional; blue: the proposed.)

In the *car* sequence (Fig.4), a white car is driving on a crooked road which is covered by snow. The image size of each frame is 768x288 pixels. The reference target model is generated from the manually located initial car region which is represented by an ellipse with the axes of 55 and 26 respectively. The tracking results are depicted in Fig.4. In this sequence, color distribution of the target is very similar to its surrounding background, especially the upside car region. In can be easily observed, with the traditional model, the tracked region deviates from its correct location gradually by mis-including the similar background. At the frame 175, the target object is lost while a tree is captured instead. However, the proposed model is robust against the similar background confusion. This due to the fact that by introducing a transition region which has negative contribution in the estimation of the proposed color model, only unique features which are distinguishable from background are used to represent the object. Consequently the car can be exactly tracked in each frame.

5. CONCLUSION

In this paper, a novel color-based target representation solution is proposed. The proposed color model is capable of capturing the distinct features of the target object by introducing a transition region which contains the surrounding background of the object. Thus the target object is represented by the color distribution estimated from both the object region and the transition region. The including of the background information helps to reduce the possibility of mis-tracking of the similar appeared background as well as the object's sub-regions. Experiments suggest that the proposed solution provides a satisfying performance in the scenarios where the background or the object's sub-regions are similar to that of the target, outperforming the state-of-the-art solutions.

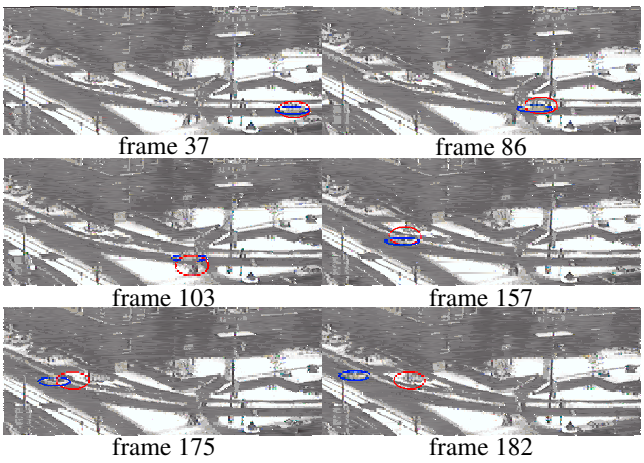


Fig. 4. Comparison of the car tracking performance of two models. (red: traditional model; blue: the proposed model.)

6. REFERENCES

[1] L. Wang S. Maybank W. Hu, T. Tan, "A survey on visual surveillance of object motion and behaviors," *IEEE Trans. Syst., Man, Cybern. C*, vol. 34, pp. 334–352, 2004.

[2] V. Ramesh D. Comaniciu and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, pp. 564–577, 2003.

[3] P. Meer D. Comaniciu, "Mean shift analysis and applications," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1197–1203.

[4] L.Gool, K. Nummiaro, K. Koller-Meier, "A color-based particle filter," in *Proceedings of the First International Workshop on Generative-Model-Based Vision*, 2002, vol. 1, pp. 53–60.

[5] A. Blake M. Isard, "Condensation - conditional density propagation for visual tracking," *International Journal on Computer Vision*, vol. 29, pp. 5–28, 1998.

[6] A.E.C. Pece P. Li, T. Zhang, "Visual contour tracking based on particle filters," *Image and Vision Computing*, vol. 21, pp. 111–123, 2003.

[7] D. Liang X. Fan, C. Qi and H. Huang, "Probabilistic contour extraction using hierarchical shape representation," in *Proceedings of the International Conference on Computer Vision*, 2005.

[8] S. Birchfield, "Elliptical head tracking using intensity gradients and color histograms," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, 1998, pp. 232–237.

[9] A. Kak J. Gao, "Multi-frame based motion estimation for semantic object tracking in the presence of occlusion," in *Proceedings of International Conference on Image Processing*, 2002, vol. 3, pp. 881–884.

[10] ftp://ftp.ira.uka.de/pub/vid-text/image_sequences/dtneu_winter/dtneu_winter.zip.