The Weight Selection in WFR Tracking Framework

YU ZHAO, HANQING LU
National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
P.O. BOX 2728, Beijing, 100080
P.R.CHINA

Abstract: In this paper, a novel method of object tracking, the Weighted Feature Representation Framework (WFR), is proposed. The basic idea of WFR is that each pixel of the target region is given different weight according to the importance of their roles in forming and influencing the features, such as position, color (or texture), motion, shape, and so on, and pixels that provide information with greater robustness for tracking have higher weight than others. In each frame, the robust feature will be different because of the different background. Thus, the feature that can be distinguished from background will be selected in each frame. In this way, the tracking precision is improved effectively at the cost of reasonable calculation.

Key-Words: Weighted Feature Representation Framework(WFR), Object tracking, feature, similarity function, robustness

1 Introduction

Object tracking can be described as: given a target region in an image, to find the new position of the region in the following images of the sequence by model match. It can be achieved by tracking the boundary or contour of the target region, or the interior of the region, or both.

The procedure of objects tracking can be divided into two steps: the feature selection and representation of the target, and the maximization of the similarity function through which the target region is located. Some pre-processing stages can also be added to improve the tracking precision, such as the color space transformation, image de-noising, and so on

In all these tracking methods, the model match plays a key role. In general, the model can be classified into the three main categories of increasing model complexity: region- or blob-based model; 2D appearance model of an object, and an articulate 3D model of an object.

In the 2D appearance model match, the classic method is Baumberg's[7] and Haritaoglu's[8]. Baumberg's *Leeds People Tracker* [7] has an adaptive shape model and occlusion reasoning, so their method has a good level of detection and tracking robustness. But their system only models shapes of the walkers, with a sufficiently large part of the body outline visible. This means that the tracker cannot be used to detect and track sitting people, and difficulties arise when tracking groups of people. Haritaoglu [8] builds *the W^4 system* to analyze *what* people are doing, *where* and *when* they do it and *who* is doing it. The main idea is to interpret and track silhouettes of foreground blobs using a feature-based approach. The way the system models the appearance of a human being in the image is very general, it has high versatility and the system is integrated with behavior analysis routines to detect simple interactions between people. But the W^4 system needs a very good and robust motion detector.

Gavrila and Davis[10], Sidenbladh[9] use an articulate 3D model of the object. The more detailed the model for object detection and tracking, the better the system can handle the particular situations for which it is trained. However, systems with complex 3D models, e.g. Sidenbladh's 3D People Tracker [9], are currently too slow to be adopted in real-time systems. And they also require a special camera/scene setup, such as the Gavrila's 3D Model-based People Tracker in [10].

The region- and blob-based tracking methods have some advantages compared with the 2D and 3D models. For example, Comaniciu [1], [2], [4] use the mean shift, a nonparametric statistical method, to solve tracking problems. It proves discrete data the convergence of a recursive mean-shift procedure to the nearest stationary point of the underlying density function. Due to the simplicity of the model, the speed of the algorithm is higher than that obtained by 2D or 3D models. But the algorithm has the low precision in some cases, e.g. the background and the foreground have similar features.

In this paper, we propose a novel method for objects tracking, the weighted feature representation Framework (WFR). As we know, for the target

region, the positions, colors, motions of its different parts play different roles in the tracking, which is closely related to the type of the object being tracked. While in many tracking algorithms such features have the same or undistinguished weight. So our approach gives the variable weight for the features with different roles.

The outline of the paper is as follows: Section 2 presents the concept of the weighted feature representation framework and the application of the WFR framework in the tracking of human bodies. Some experiments have been shown in the end of this section. Section 3 presents a summary about the WFR framework.

2 The WFR Framework

In tracking problems, one object in different frames will probably experience changes in its appearance or shape, so how can we track it? In fact, there are always the most characteristic and stable features that identify the object. So we should focus on the most important and robust features of the object instead of treating equally every pixel in the image. The WFR algorithm is aimed at such objectives, and next we will discuss the WFR algorithm in detail.

2.1 The Concept of WFR Framework

To characterize the target region, a feature space is chosen. And then the representation of the target model and the target candidate is established according to the feature space. The representation approach has an important effect on obtaining the correct target region in the new images. The importance of each feature of the target region can be analyzed and the weight function to the region can be established. The weight function should satisfy such condition: the more robust the feature is, the higher its weight is.

For features that describe the position of different localities in the target, the region center should have higher weight than the region boundary, since the center has higher stability. For example, when there is partial occlusion in the object, there will be a higher probability for boundary occlusion than the center occlusion. But in some special target tracking, the object with different parts should be divided into several parts according to the importance of each part, e.g. the human body can be three parts: head, torso and bottom, in each part, the center weight will higher than the boundary one.

For the color feature, we can find some main colors in the target region by the color histogram or other descriptions. The primary color components should have higher weight than the minor ones. For instance, in the human body tracking, the head, torso and bottom color should be the important feature; especially the torso color can be the key feature to distinguish the target from the other objects. We also select the main colors which can be distinguished obviously from the background, since the human eyes can find out the objects from the background by the special color that is easy to distinguish from the background.

For the motion feature of human being, each part of the body moves in different directions. For instance, when a man walks through a camera scene, the head can rotate left and right, the arms and the legs can move up and down. But there is a main motion that we should concentrate: the mean motion of every parts of the body. So the motion similar to the main motion of the object should have higher weight than others.

For the shape feature, different objects have different characters, which distinguish the objects from others, so it should have higher weight. The shape feature is related to the tracked object, e.g. in the car images, the straight line or parallel edge feature should be more important, while the ellipse shape is more concerned in the head tracking of human being. And in the shape analysis, we have to add the motion feature to it, since we are only concerned with the objects of similar motions.

All weight features can be synthesized into one formula: $\hat{I}(x) = \prod_{i=1}^{n} W_i(x) \cdot I(x)$

where $\hat{I}(x)$ is the weighted image function, $W_i(x)$ is the weight of each feature, I(x) is the original image, and n is the number of features that are used to calculate the weight.

2.2 The Application of WFR in the Tracking of Human Body

In human body tracking, we should analyze several features as follows: color, position, motion and shape. The weighted image function can be written as:

$$\begin{aligned}\hat{I}(x) &= \prod_{i=1}^{n} W_i(x) \cdot I(x) \\ &= W_{position}(x) \cdot W_{color}(x) \cdot W_{motion}(x) \cdot W_{shape}(x) \cdot I(x)\end{aligned}$$

2.2.1 Position Feature

The Position of human body should be divided into three parts, head, torso, and bottom. In the center of each part, the weight can be high, so we first get the center, radius and percent of each part and use the Gaussian Kernel weight function, the weight of each pixel can be written as:

$$w^i_{position} = \frac{1}{s_a s_b s_c} e^{-\frac{1}{2}(\frac{a_i - a_0}{s_a})^2} e^{-\frac{1}{2}(\frac{b_i - b_0}{s_b})^2} e^{-\frac{1}{2}(\frac{c_i - c_0}{s_c})^2}$$

where a, b, c is each pixel value of the three parts, a_0, b_0, c_0 is the center. s_a s_b s_c is the bandwidth (scale) of the three parts. The sketch map is shown in Fig.1

2.2.2 Color Feature

As for color feature, we first change RGB color space into another form:

$$r = \frac{R}{R+G+B},\ g = \frac{G}{R+G+B},\ s = \frac{R+G+B}{3}$$

Then we rescale the color component into 0~n-1 (n is the valid range of the new color). And the weight function is:

$$w_{color}(x_i) = \frac{1}{n} e^{-\frac{1}{2}(\frac{b(x_i) - m_0}{s_0})^2} \cdot \sum_{i=1}^{n} d[b(x_i) - u]$$

where the function $b: R^2 \rightarrow \{1...m\}$ associates to the pixel at location x_i, d is the Kronecker delta function. m_0 and s_0 is the mean and deviation of the target model. m_0 gives a mean color of the target region, and the color value near m_0 will have high weight. The sketch map is shown in Fig.2. The color value represents the weight, and we can see the target region have higher weight than other regions.

(a) (b) (c) (d)

figure 1. the weight position feature. (a) the original image. (b) the target region image. (c) the weight function. (d) the weight image

2.2.3 Motion Feature

Human being motion is very complex. But in the human body tracking, there is a main motion of the body, and other motions like hand, head, and so on, aren't what we concern. We can calculate the motion of each pixel m_i and the main motion is the mean of the motions:

$$m_0 = \frac{1}{n}\sum_{i=1}^{n} m_i \cdot$$

And the weight function can be written as follows: $w^i_{motion} = k(\| \frac{m_i - m_0}{h} \|)$

The motion feature will be used in object locate, we can estimate the next location of the target region according to the motion analysis.

2.2.4 Shape Feature

Human body is a non-rigid object, but we can divide the body into several rigid parts. The head is similar to an ellipse, thus we only search the ellipse edge in the image, and calculate its parameter. The torso and the bottom parts are relevant to the clothes of the man, so we have to learn the shape of the parts at first. However, the human body is similar to a rectangle region. The motion feature can be incorporated into the shape feature, and we can detect the rectangle motion (that is the rectangle region with identical motion) in the image to associate the torso and the bottom parts.

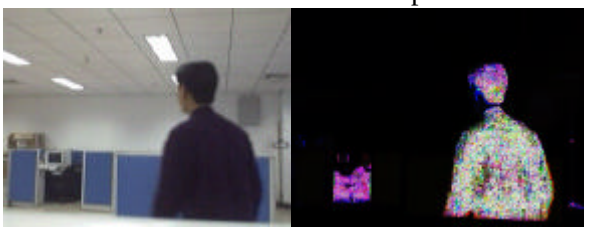

figure 2. the weight color feature. (a) the original image. (b) the color weight image

2.2.5 Tracking

Let I_0 represent the target model, I_1 represent the target candidate after WFR processing. We use the Bhattacharyya Coefficient as the similarity function, the distance between two distributions as:

$$d(y) = \sqrt{1 - r(I_0, I_1)}$$

where we chose

$$r(I_0, I_1) = \sum_{u=1}^{n} \sqrt{I_0(u) \cdot I_1(u)}$$

the sample estimate of the Bhattacharyya coefficient between I_0 and I_1, u is the region bit.

The tracking algorithm is presented below:

(Bhattacharyya Coefficient Maximization)

Given:

The target model I_0 and its location y_0 in the previous frame.

1. Initialize the location of the target in the current frame with y_0
2. Calculate the weighted image $I_1(y_0)$ using the WFR algorithm. And evaluate
$$r(I_0, I_1(y_0)) = \sum_{u=1}^{n} \sqrt{I_0(u) \cdot I_1(u, y_0)}$$
3. Estimate the main motion of the target candidate and calculate the next candidate center of the target region according to the mean motion m_0.
4. Compute $r(I_0, I_1(y_1)) = \sum_{u=1}^{n} \sqrt{I_0(u) \cdot I_1(u, y_1)}$

5. While $r(I_0, I_1(y_1)) < r(I_0, I_1(y_0))$
   Do $y_1 \leftarrow \frac{1}{2}(y_0 + y_1)$ and evaluate $r(I_0, I_1(y_1))$
6. If $\|y_1 - y_0\| < e$ stop.
   Otherwise set $y_0 \leftarrow y_1$ and go to step 2

If the target model gradually becomes invalid when the distance between the target model and the target candidate increases, we have to update the target model according to the current target region. The Fig. 3 shows some experiments of image sequences.

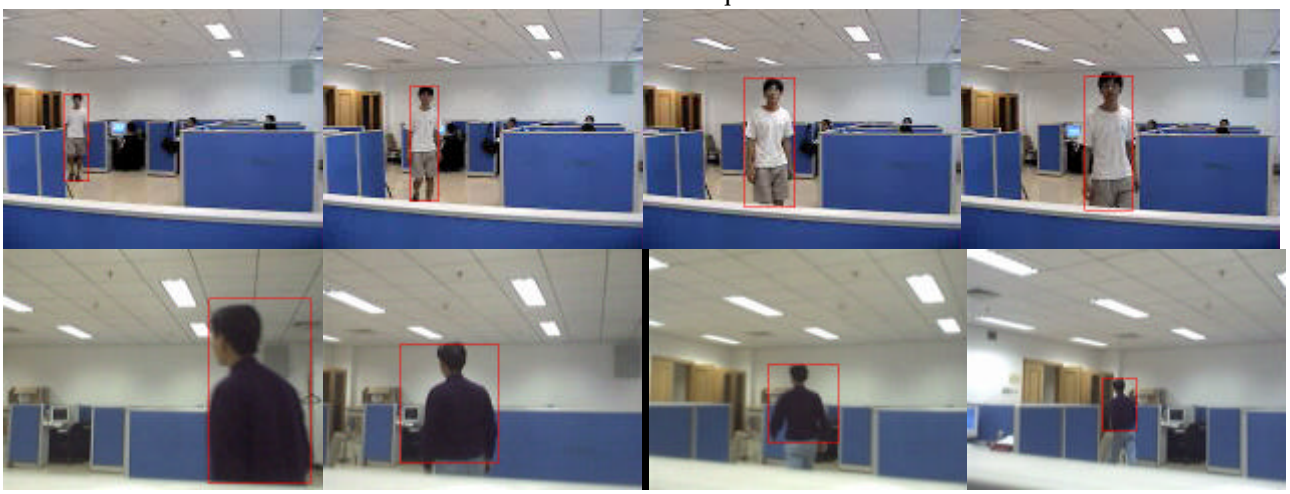

figure 3. the tracking result of two image sequences. Top: the image sequences that the man walks up to the camera. Bottom: the image sequences that the man walks far from the camera.

3 Summary

The WFR algorithm presents a novel framework in the objects tracking, which focuses on the analysis of the importance and robustness of target region. It weights the target region according to the importance and robustness of the target region feature. The algorithm has been implemented in some image sequences. It performs well as the Fig.3 shows. Though the WFR algorithm uses many features for processing, the computation procedures are not complex. So the speed of the tracking is acceptable. And the features are combined together to give weight to different pixels. For example, in the shape analysis of the human body, we should process the motion feature and the shape feature at the same time.

In the future work, what we should do is to improve the weight function to solve the multi-object tracking and to enhance the robustness of the tracking in the realtime system.

References:

[1] Comaniciu, D., Ramesh, V. and Meer, P., *"Kernel-Based Object Tracking"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 25, No. 4, April 2003

[2] Comaniciu, D., Ramesh, V. and Meer, P., *" Real-Time Tracking of Non-Rigid Objects using Mean Shift,"* IEEE Conference on Computer Vision and Pattern Recognition, Vol II, 2000, pp. 142-149

[3] Robert T. Collins, *"Mean-shift Blob Tracking through Scale Space"*, IEEE Conference on Computer Vision and Pattern Recognition Vol II, 2003.

[4] Dorin Comaniciu, Peter Meer, *"Mean Shift: A Robust Approach Toward Feature Space Analysis"* , IEEE Transactions on Pattern Analysis and Machine Intelligence 24(5): 603-619, 2002.

[5] A. Elgammal, R. Duraiswami, and L. S. Davis *" Probabilistic Tracking in Joint Feature-Spatial Spaces"* IEEE – Conference on Computer Vision and Pattern Recognition (CVPR 03), Madison, Wisconsin, June 16-22, 2003.

[6] A. Elgammal and L. S. Davis, *"Probabilistic Framework for Segmenting People Under Occlusion"*, The Eighth IEEE International Conference on Computer Vision, Vancouver, Canada July 9-12, 2001

[7]A. Baumberg and D. Hogg, *" An Efficient Method for Contour Tracking using Active Shape Models"*, in proc. IEEE Workshop on Motion of Non-rigid and Articulated Objects, papers 194-199, 1994

[8] I. Haritaoglu, D. Harwood, and L. Davis, *" w4s: Areal-time system for detecting and tracking people in 2.5d,"* in Computer Vision, ECCV, 1998.

[9] Sidenbladh, H. *" Probabilistic Tracking and Reconstruction of 3D Human Motion in Monocular Video Sequences"*, Ph.D. thesis, KTH, Sweden. TRITA.NA-0114, 2001.

[10] D. Gavrila and L.S.Davis, *"3-D model-based tracking of humans in action: a multi-view approach"*, CVPR 1996:73-80, 1996.