Tracking Multiple Circular Objects In Video Using Helmholtz Principle

Snehasis Mukherjee and Dipti Prasad Mukherjee
Electronics and Communication Sciences Unit
Indian Statistical Institute
Kolkata, India
{snehasis_r, dipti}@isical.ac.in

Abstract— A novel algorithm is introduced to track multiple circular objects present in a video using Helmholtz perception principle. First, segmentation of circular objects in the video frame is performed using the perception principle and then same perception principle is applied to track the circular objects. For each circular object present in video, we have taken an assessment of the meaningfulness of the shift of its center of gravity and meaningfulness of the deviation of the direction of movement of the object due to inter-frame displacement. We have shown that a logical threshold in the meaningfulness value tracks circular objects in a video effectively and efficiently.

I. INTRODUCTION

Recently Gestalt hypotheses are being used for solving several problems of computer vision [1]. In most cases Gestalt theory is used to find meaningful image segment or geometric shape in an image after extracting the image level lines, which are the edges of iso-intensity surfaces present in an image [1,2,3,4]. The term meaningful here refers to the objects present in an image which exhibit some specific features (such as linearity, circularity, rectangularity, etc.) of our interest. Gestalt hypotheses has also been used for detecting circular level lines [3] using the concept of number of false alarms (NFA), where the term false refers to a non-meaningful event. The NFA is the expected number of occurrences of an event, where an event is defined as the occurrences of the features of some pattern in an image. Our contribution in particular is to extend the concept of NFA for tracking multiple circular objects in a video.

Here we have estimated meaningful displacements of the circular objects present in any two consecutive frames using the Helmholtz perception principle. Helmholtz principle for images states that every large deviation from randomness in an image should be perceptible, provided, the large deviation corresponds to an a priori fixed list of geometric structures (such as, straight line, circle, rectangle, etc.) [1]. In this paper we have extended this idea to the context of video as, large deviations in any feature (such as displacement of the center of gravity of an object in two consecutive frames) modulo a perceptible limit is meaningful in case of a successful tracking.

II. BACKGROUND

We define the number of false alarms (NFA) as the expected number of occurrences of an event, where an event is defined as the occurrences of the features of some pattern such as, smoothness, circularity, rectangularity, etc. in images or inter-frame displacement of an object in video. For an exhaustive trial, if the NFA is sufficiently small, then the event is meaningful, i.e., the event cannot occur out of a uniform random process [1]. Then we need a threshold to indicate the value $\varepsilon$ below which, we consider the NFA as...
sufficiently small and hence the event is meaningful. In that
case, the event is called as an $\varepsilon$-meaningful event.

Now, we present briefly the method of using the
concept of $\varepsilon$-meaningfulness to detect the circular objects
in an image. For details of this methodology, the reader may
refer to the paper [3].

Assume an image $I \in Z^2$ with intensity levels $[0, I]$. For
each closed intensity level line $C$ of $I$, first sample the
level line at sample length $\delta$ apart to get $(n+1)$ number
of points on the curve, say $p_1, p_2, p_3, \ldots, p_n$. The criterion for
any level line $C$ to be circular is that the differences
between the distances of any two consecutive Sample points
on $C$ from the CG of $C$, say, $d_{i-1}$ and $d_i$, are sufficiently
small for each of the points $p_i$ on $C$. So, for each $i$ in
$0 \leq i \leq n$ , calculate $k_i = d_i - d_{i-1}$. For $i = 0$, $d_{-1}$ is the
distance of $p_n$ from the CG of $C$. For a circular object, $k_i$
should be close to zero. Clearly, $|k_i|$ achieves its maximum
possible value $\delta$ when the two consecutive points $p_{i-1}$ and
$p_i$ and the CG are collinear. The $|k_i|$ values are independent
to each other and therefore, i.i.d. sequence of random variables
with uniform distribution in $[0, \delta]$ . In reality, the maximum allowable value of $|k_i|$ , say, $k_{\max}$
should be far less than $\delta$ for the level line to be circular.
Let $N_C$ be the number of closed level lines obtained for the
level $l$ for which one level line $C$ has been obtained. Then, in
this case NFA can be defined as follows:

$$NFA(C) = N_C(\max_{\beta} \{k_{\max} / \delta\} ) \text{ for } |k_i| < k_{\max} \text{ and }$$

$$NFA(C) = N_C \text{ otherwise} \tag{1}$$

For each point $p_i$, calculate the NFA. As mentioned
earlier that if NFA given by (1) is less than $\varepsilon$ up to the
point $p_i$, then the level line is $\varepsilon$-meaningful up to the point
$p_i$. If for all the points $p_i$, $0 \leq i \leq n$, the level line becomes
$\varepsilon$-meaningful, then we conclude that, the level line is $\varepsilon$
meaningful, i.e., in this case the level line is circular. Here,
the value of $\varepsilon$ has been taken as 1, as it is convenient to
assume that an event cannot be the outcome of a uniform
random process if the expected number of occurrences of the
event is less than 1.

In the next section, we use the concept of NFA and the
method of detecting circular objects to track multiple circular
objects in a video.

III. TRACKING CIRCULAR OBJECTS IN VIDEO

Let us consider any two consecutive frames in a video.
We have first detected the circular objects in each of the
frames by calculating NFA from (1) and using the method
described in Section 2. So, our next task is to find out the
correspondence between the detected circular objects of one
frame with that of the other. In order to do so, as noted in
the introduction, we have compared the distances traversed
by each cell of the previous frame with respect to each of
the cells detected in the current frame. The correspondence
between two circular objects of the current and previous
frames is established if the distance traversed is meaningful
compared to the other possible distances traversed by a cell
of the previous frame.

Let $o_1, o_2, \ldots, o_m$ be the $m$ objects detected in the
previous frame, and $q_1, q_2, q_3, \ldots, q_t$ be the $t$ objects
detected in the second frame. For each given $q_j$, $1 \leq j \leq t$
, we calculate displacements $\beta_y$ of the CG of $o_i$, $1 \leq i \leq m$
from the CG of the object $q_j$. If the video frames be of size
$M \times N$ , then $\beta_y$ can get a maximum value of $\sqrt{M^2+N^2}$
which is the distance between any two diagonally opposite
corner points of the frame. So, as the $\beta_y$ values are
independent to each other, $\beta_y$ are i.i.d. sequence of random
variables with uniform distribution in $[0, \sqrt{M^2+N^2}]$ . In
reality, the maximum allowable value of $\beta_y$, say $\beta_{\max}$
should be close to zero for $q_j$ of the second frame to be
same object as $o_i$ of the first frame.

Now our next task is to define the NFA for this
problem. As discussed in the Section 2, NFA is the expected
number of occurrences of the event such that $q_j$ is the
displaced position of $o_i$ of the previous frame. So, we first
have to calculate the probability $P(\beta)$ of the event. Since
$\beta_y$ is uniformly distributed over $[0, \sqrt{M^2+N^2}]$ , so the
probability is given by,

$$P(\beta) = \beta_{\max} \sqrt{M^2+N^2} \tag{2}$$

Now, since the total number of objects in the first frame is
$m$, and since we are doing our experiment for all the $m$
objects, we can define the NFA using the probability given
by (2) as follows:

$$NFA = m P(\beta) \text{ when } \beta_y < \beta_{\max} \tag{3}$$

$$NFA = m \text{ otherwise}$$

If the NFA given by (3) for $o_i$ is less than a given number
$\varepsilon$ , then the above said event is said to be $\varepsilon$-meaningful,
i.e., a given $q_j$ in the current frame can be recognized as
the same object $o_i$ of the previous frame. To avoid conflict,
we start with a sufficiently small value of $\beta_{\max}$ and then
update its value by $\beta_{ij}$ if $q_j$ and $o_i$ be detected as the same object.

Besides assessing the meaningfulness of the target displacement, we have also introduced a measurement of meaningfulness of the direction of movement of the target objects. Consider ($f$-2), ($f$-1) and $f$th consecutive frames of the video. Also, assume the correspondences of targets between frames ($f$-2) and ($f$-1) are already established. For each detected target in $f$th frame, let us define $x(j_1, i_{1,1})$ as the slope of the tracking path connecting $i$th target in $f$th frame with $i'$th target in ($f$-1)th frame. Similarly, $x(i_{1,2}, i_{2,1})$ is the slope of the tracking path of $i'$th target between frames ($f$-2) and ($f$-1). Then the absolute difference in the tracking path direction for consecutive frames is given by,

$$\theta_{ij} = \tan^{-1}(x(j_1, i_{1,1})) - \tan^{-1}(x(i_{1,2}, i_{2,1})).$$  

(4)

The value of $\theta_{ij}$ should be close to zero and clearly, a i.i.d. sequence of random variables defined on $[0, \pi]$. Then, if $\theta_{ij}$ is the maximum allowable value of $\theta_{ij}$ for $o_i$ to be the same object as $q_j$, then, the probability that the direction of the object to be meaningful is given by,

$$P(\theta) = \left(\frac{\theta_{ij}}{\pi}\right).$$  

(5)

Then, we define NFA of the tracking path directional consistency with the similar argument as (3) as follows:

$$NFA = m P(\theta) \text{ when } \theta_{ij} < \theta_{ij}.$$

(6)

$$NFA = m \text{ otherwise}$$

The NFA given by (6) should also be less than $\epsilon$ for the above said event to be meaningful.

Therefore, to integrate meaningfulness of both tracking displacement and direction, we have done an AND operation between NFAs given by (3) and (6). If the result be less than $\epsilon$ then we may conclude that $o_i$ and $q_j$ are the same objects in the two consecutive frames. Now, the problem is to choose a value of $\epsilon$ for this integrated framework. Note that if the NFA of an event is less than $\epsilon$ given by (3) and (6) separately, then the NFA after integrating (3) and (6) through AND operation must be less than $\epsilon$ as well. Therefore, $\epsilon$ can be chosen as 1 for the event to be meaningful satisfying both (3) and (6). The algorithmic steps are detailed below.

**Algorithm:**

**Step 1:** Fix the value of $\beta_{ij}$, $\theta_{ij}$ and $\epsilon$.

**Step 2:** Detect all the circular objects present in the first frame using NFA given by (1).

**Step 3:** Find the CG of all the objects.

**Step 4:** For frame=2 to end do

**Step 4(a):** Detect all the circular objects present in the frame using NFA given by (1).

**Step 4(b):** Find the CG of all objects in the current frame.

**Step 4(c):** For CG of each object in this frame do

**Step 4(c)(i):** For CG of each object in previous frame do

**Step 4(c)(ii)(I):** Find the distance $\beta_{ij}$ between the two CG.

**Step 4(c)(ii)(II):** Find the absolute difference $\theta_{ij}$ using (4) starting 3rd frame.

**Step 4(c)(ii)(III):** If $NFA < \epsilon$

- then the two objects may be same and continue with the next object of the previous frame; update $\beta_{max}$ by $\beta_{ij}$.
- else the two objects are not same and continue with the next object of the previous frame.

**Step 5:** The remaining objects in the present frame are new objects, so color them by another color.

In the next section, we show the results of our experiments with this methodology.

IV. RESULTS AND DISCUSSIONS

We have tested our method on videos of blood cells moving inside the veins of mouse cremaster muscle. A typical sequence of consecutive frames is shown in Fig. 1(a) and (b). The corresponding cells are shown using same color in both the frames.

We have already discussed about the value of $\epsilon$. The parameter $\beta_{max}$ is set based on the domain knowledge. Since, the cell movements captured at 30 frames/second are not significant, $\beta_{max}$ value is set at 5 pixels. For the same reason and also since, the blood cells do not show erratic movements, the other parameter $\theta_{max}$ is set as 0.5 radian.

We have tested the proposed approach on 30 consecutive frames of the video of moving blood cells. It has effectively detected nearly 70% of the white blood cells present in the frames. In only 2 cases in the frame #778 in Fig. 1(a), false positives are detected but those were not tracked. However, 80% of the detected targets were tracked effectively upto 30 frames.

Fig. 1(a): Frame #778. (b): Frame #779
The ROC curve given in Fig. 2 shows the efficiency of our proposed algorithm. The area under the ROC curve is calculated as 0.8035 which is satisfactory (ideal value should be 1).

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2.png}
\caption{ROC Curve}
\end{figure}

A tracking result is shown graphically in Fig. 3 plotting the x and y ordinates of the CG of 3 moving blood cells in the 30 consecutive frames. Different cell paths are clustered separately and three tracks are clearly separable. Due to temporary occlusion, some small fragmented track paths are also detected.

In comparison to approaches like Hough Transform, the proposed approach is computationally far superior. While Hough Transform requires fixing of similar number of parameters, it requires at least three $O(n^7)$ operations. In contrast, the current approach works at approximately $O(lp)$ where $l$ is the number of level lines and $p$ is the number of sampled points on the level line. Both $l$ and $p$ are significantly smaller compared to image and accumulator space dimensions of Hough transform.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig3.png}
\caption{Blood cell positions in consecutive 30 frames of the video. For details, refer to the text.}
\end{figure}

V. CONCLUSIONS

We have introduced a successful tracking framework that requires parameters which are problem specific and their selection is straightforward. The tracking performance is significant given that most of the circular objects present even in the frames of noisy video could be detected and tracked. We have shown that the objective assessment of meaningfulness works both for segmentation and tracking. Our future plan is to track occluded objects in a video, establish a mathematical model to find the values of the parameters and extend the framework towards a generalized shape detection and tracking.

REFERENCES


