BTP REPORT
on
Fragment Based Object Tracking

by
Utkarsh Kumar Shah
B.Tech, 4th year
Y6510
Guide: Prof. A. Mukerjee
Dept of Computer Science and Engineering
IIT Kanpur - 208 016
INDIA

November 11, 2009

1 OBJECT TRACKING

Object tracking is a method of following an object through successive image frames to determine its relative movement with respect to other objects. Video tracking is performing object tracking in a video. An algorithm analyzes the video frames and outputs the location of moving targets within the video frame. Unknown objects in an image are often called blobs. And if we classify them according to our interest, it becomes an object, such as a vehicle or person. Hence, in this way blobs are classified as objects during image analysis.

1.1 Objective

There are various existing algorithms for object tracking. Among which the widely used one is mean shift tracking which I will be describing later.The other techniques used so far includes blob tracking (segmentation of object interior), contour tracking (detection of object boundary - condensation algorithm), visual feature matching, levelset tracking and lot more.

Many existing algorithms, segment each video frame to determine the objects. There are algorithms that derive the objects based on the motion between frames. But these tracking algorithms are limited and not able to handle some complex situations such as new tracks (object starts moving), ceased tracks (object stops moving) and possible collisions (objects move together).There are lot of other challenges in object tracking which I will be describing further in report.

Fragments based level set tracking is one of the robust object tracking algorithm. It handles decently challenging situations such as non-rigid deformations, rapid motion, full/partial/self occlusion and multiple object tracking. So, I am aiming to implement this algorithm.

2 CHALLENGES IN OBJECT TRACKING

2.1 Occlusion

Occlusions can be classified as either static or dynamic based on the respective nature of the occluder i.e. occlusions caused by either the static scene structure or other moving objects

2.1.1 Static occlusion

Occlusion caused by a static structure like tree, pillar etc.

Figure 1: Static Occlusion: Occlusion caused by a tree

2.1.2 Dynamic occlusion

Occlusion caused by moving objects like two cars moving in opposite direction, one hiding the other for short interval of time. It can be further classified into merging and splitting of various objects. For example, two persons coming along and then going in separate ways or vice versa.

Figure 2: Dynamic Occlusion: moving grey car is occluded by another moving white car

Figure 3: Splitting of objects

Occlusion can be full or partial. It depends on the size of the object which occludes.

2.2 Object Appearance in Image

Object appearance includes its shape and visibility features such as color distributions, gradient information, high curvature points etc. These can create problem due to sudden changes in any of these features. Also, the object's shape might change as in case of a car coming towards the camera and then suddenly change the direction. Change of shape can be within object itself like opening and closing of car's door, a running animal etc.

Figure 4: A running monkey and two views of plate: shows how object shape can change frequently

2.3 Real-time Constraint

With a lot of time given, an exhaustive search would be able to locate an object in the image frame properly. However, in most practical cases, we need to use only a small portion of the model space to reduce the computational burden. In an offline processing, the data content can be very large which limits use of any number of object models to achieve the desired results in real time.

3 Work Done So Far

3.1 Background Subtraction

Background subtraction is the starting step for object tracking. We (Sudhamsh and I) worked together on single gaussian background subtraction algorithm. We tested it on PETS2000 dataset.

Figure 5: The original image on left and the the output after running our background subtraction code

Then I read various papers in object tracking algorithms. Of which I am mentioning the few algorithms such as Mean shift tracking algorithm further.

4 OBJECT TRACKING ALGORITHMS

4.1 Mean Shift Tracking Algorithm

[2, 4, 7] In this algorithm moving objects are represented by their color-histograms. Therefore the object are tracked by histogram estimation. It is an iterative algorithm which compares the histogram of the original object in the current frame and that of candidate regions in the next frame of image. The aim is to find maximum correlation between the two histograms.

4.1.1 Object Model

A reference target model is chosen and represented by its probability distribution function in color space. Let x_i, i = 1..n, denote pixel locations of model centered at 0. Color distribution is represented by discrete m-bin color histogram. Let b(x_i) denote the color bin of the color at x_i. Then, the probability p of color u in the model is:

q_u = C$\sum_{i=1}^{n}$ k($\|x_i\|^2$) δ(b(x_i)-u)

where C is the normalization constant

C = $[\sum_{i=1}^{n}$k($\|x_i\|^2$)$]^{-1}$

Kernel profile k weights contribution by distance to centroid.

δ is the Kroneckar delta function. That is, contribute k($\|x_i\|^2$) to q_u if b(x_i) = u

4.1.2 Target Candidate

Let y_i, i = 1..m, denote pixel locations of target centered at y. Then, the probability p of color u in the target is

p_u(y) = C_h $\sum_{i=1}^{m}$ k($\|\frac{y-y_i}{h}\|^2$) δ(b(y_i)-u)

where C_h is the normalization constant

C_h = $[\sum_{i=1}^{m}$k($\|\frac{y-y_i}{h}\|^2$)$]^{-1}$

Figure 6: [7] The object model is centered at 0 and the target candidate is centered at y

4.1.3 Color Density Matching

Using Bhattacharya coefficient ρ

where ρ(p(y), q) = $\sum_{i=1}^{m} \sqrt{p_u(y) q_u}$

ρ is the cosine of vectors $(\sqrt{p_1},..\sqrt{p_m})^T$ and $(\sqrt{q_1},..\sqrt{q_m})^T$. Large ρ means good color match. For each image frame, find y that maximizes ρ. This y is the location of the target

4.1.4 Tracking Algorithm

Given q_u of model and location y of target in previous frame:

1. Initialize location of target in current frame as y
2. Compute p_u(y) and ρ(p(y), q)
3. Apply mean shift: Compute new location z as
   z = $\frac{\sum_{i=1}^{m} k(\|\frac{y-y_i}{h}\|^2) y_i}{\sum_{i=1}^{m} k(\|\frac{y-y_i}{h}\|^2)}$
4. Compute p_u(z) and ρ(p(z), q)
5. While ρ(p(z), q) ¡ ρ(p(y), q), do z $\leftarrow$ 1/2(y+z) (to validate target's new location)
6. If $\|$z-y$\|$ is small enough, stop. Else set y $\leftarrow$ z and goto 1

4.1.5 Results

Figure 7: This is the result of mean shift tracking implemented by P. Guha in IITK. (1) shows a car marked by maroon box and a rickshaw by yellow, (2) shows that due to occlusion by car the rickshaw is not marked and the yellow box of rickshaw is not having object in it, (3) Rickshaw is identified as a new object(marked by blue box) and yellow box marks a person which earlier was marked by red box.

As can be seen in the result, track of various object gets lost midway, usually because of partial or full occlusion. It is highly needed to develop an algorithm which can keep track of these problems.

4.2 ADAPTIVE FRAGMENTS BASED TRACKING

4.2.1 Approach [1, 6]

Tracking Framework: Target and background is modeled as a mixture of Gaussians. A strength map is computed indicating the probability of each pixel belonging to the foreground.

Image Segmentation: Target is divided into multiple fragments.

Contour Extraction: Contour is extracted using level set implementation.

Update Mechanism: The fragments are automatically adapted to the image data, being selected by an efficient region-growing procedure and updated according to a weighted average of the past and present image statistics.

The extracted target boundaries are used to learn the dynamic shape of the target over time, enabling tracking to continue under total occlusion.

4.2.2 Tracking Framework

They used Bayes' rule and an assumption that the measurements are independent of each other and of the dynamical process to model the probability of the contour Γ at time t

Bayesian Formulation

$$\mathrm{p}(\Gamma_t | I_{0:t}, \Gamma_{0:t-1}) \propto \underbrace{p(I_t^+ \mid \Gamma_t)}_{target} \underbrace{p(I_t^- \mid \Gamma_t)}_{background} \underbrace{p(\Gamma_t \mid \Gamma_{0:t-1})}_{shape}$$

where Γ_t: contour at time t

$I_{0:t}$: image data of all frames, $\Gamma_{0:t-1}$: previously seen contours

I_t^+ captures the pixels inside Γ_t, I_t^- captures the pixels outside Γ_t

Object Modeling and Strength Image

Positive values in the strength image indicate pixels that are more likely to be long to the target than to the background, and vice versa for negative values.

4.2.3 Region Segmentation

Mode-seeking region growing algorithm:

do {

- Pick a seed point that is not associated to any fragment
- Grow the fragment from the seed point based on the similarity of the pixel and its neighbor's appearance

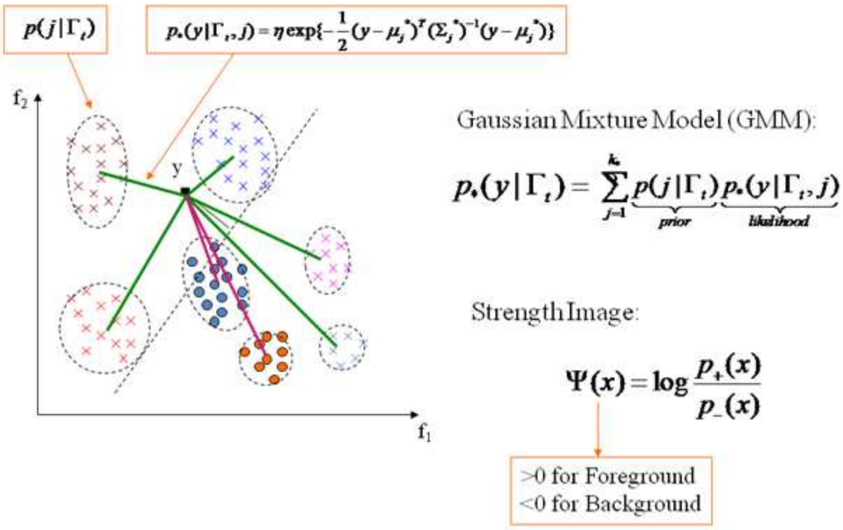


Figure 8: [1] Cross marks indicate foreground parts and circles indicate background part. The strength image is computed using log ratios of probabilities

- Stop growing the fragment if no more similar pixels are present in the neighborhood of the fragment

} until all pixels are assigned

Seed point:

Eigen values of 3x3 RGB covariance matrix

$$\eta(x)=\prod_{i=1}^{3}\lambda_i(x)$$

$$S=<v_1,\dots,v_n> \quad \text{where} \quad v_i=(x_i,y_i,\eta_i)$$

- Pick the minimum element in S. Create a region to hold the pixel and add the neighbors in a fixed window.
- Compute Mean μ_j and Covariance Σ_j of the region.
- Likelihood:

Mahalanobis distance

$$\Psi_j(x)=MD(f(x),\mathrm{N}(\mu_j,\Sigma_j))-\tau$$

Configurable parameter

- Grow the region as before with two additional steps:

  Update μ_j, and Σ_j , as a new pixel is added

  Remove the corresponding element in S if a pixel is added
- Continue above steps if S is not empty.

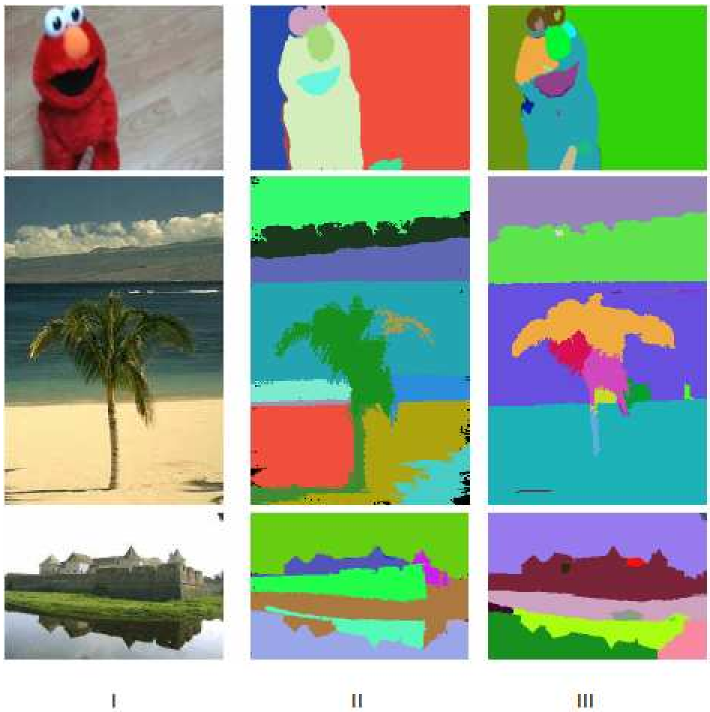


Figure 9: [1] Segmentation result of region growing algorithm compared with other methods. I. Original Image, II. Region Growing, III. Mean Shift. Mean shift approach takes roughly 30 seconds to segment an image of size 320X240. This approach completes the segmentation process in less than a second.

4.2.4 Contour Extraction

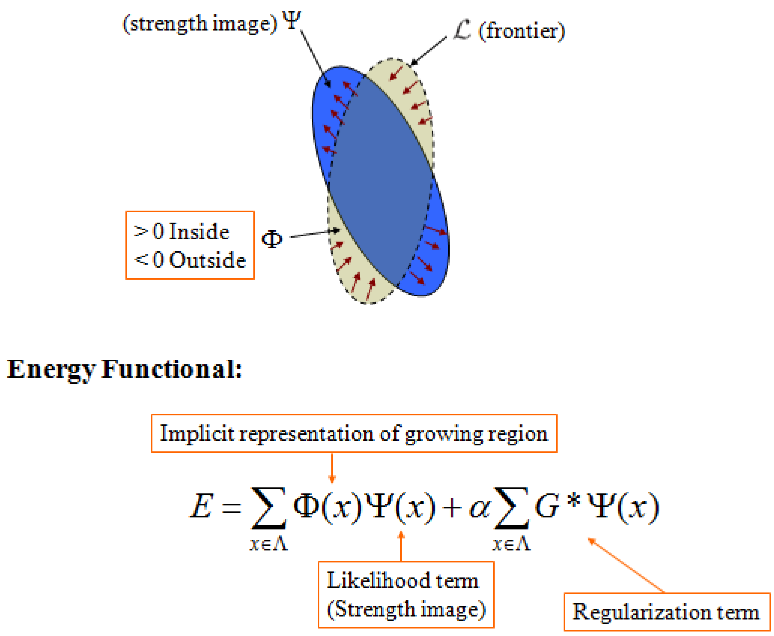
The region growing algorithm starts with a seed region,Ω,and uses two kinds of representation: an explicit representation using a singly linked list, L, and an implicit representation,ϕ, similar to the level sets approach. ϕ is initialized as follows:

ϕ(x)= { +1: x $\in$ Ω; -1: otherwise }

The singly linked list, L, is initialized as:

L = {x: x $\in$ Ω, $\exists$ x' $\in$ N_4(x) such that x' !$\in$ Ω

L represents the boundary of the region being grown



The region evolution proceeds by evolvingϕ over time to maximize the energy function given below.

Region growing algorithm

Require: The likelihood term ψ, ϕ and the list L

ϕ represents the entire region and L represents the boundary of the region

```
repeat
  for all x ∈ ℒ do
    { Expansion Step: }
    for all x' ∈ N_4(x) do
      if Φ(x') < 0 AND Ψ_g(x') > 0 then
        ℒ ← ℒ ⊕ x'                                              Expand
        Φ(x') ← +1
      end if
    end for
    if Ψ_g(x) > 0 AND (Ψ_g(x') > 0, ∀x' : x' ∈ N_4(x)) then    Remove interior points
      ℒ ← ℒ ⊖ x
    end if
    { Contraction Step: }
    for all x' ∈ N_4(x) do
      if Φ(x') > 0 AND Ψ_g(x') < 0 then
        ℒ ← ℒ ⊕ x'                                              Contract
        Φ(x') ← −1
      end if
    end for
    if Ψ_g(x) < 0 AND (Ψ_g(x') < 0, ∀x' : x' ∈ N_4(x)) then    Remove exterior points
      ℒ ← ℒ ⊖ x
    end if
  end for
until ℒ is not modified
```

Figure 10: [1] The Region Growing Algorithm

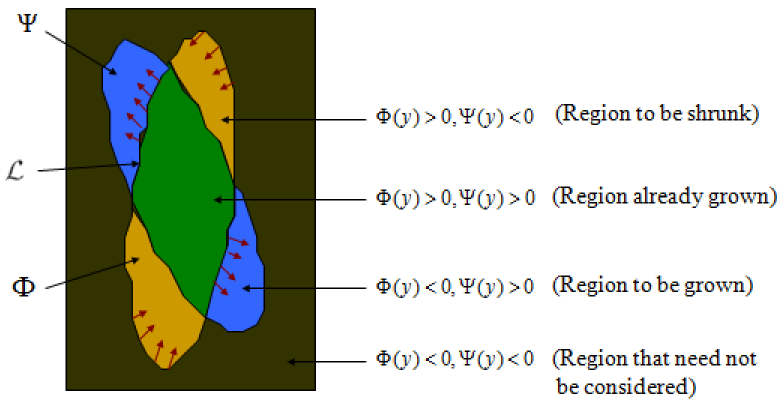


Figure 11: [1] Demonstration of the region growing algorithm where the frontier L evolves from its current position using the implicit representation ϕ and the likelihood of the region ψ.Four region types are possible using ϕ and ψ which are indicated by different colors.

4.2.5 Update Mechanism

Updating parameters of existing fragments:

Weight computed by comparing Mahalanobis distance

$$\mu_{j,t}^{*} = \alpha_j^{*}\mu_{j,0:t}^{*} + (1-\alpha_j^{*})\mu_{j,0}^{*}$$

$$\mu_{j,0:t}^{*} = \frac{\sum_{\tau=0}^{t} e^{-\lambda(t-\tau)}\mu_{j,\tau}^{*}}{\sum_{\tau=0}^{t} e^{-\lambda(t-\tau)}}$$

(function of past and current values)

Initial Model

Occluded fragments: If a fragment is associated with less than 0.2% of the image pixels, then the fragment is declared as occluded.

4.2.6 Results

Figure 12: [1] The target is tracked properly despite shape deformation and large unpredictable motion

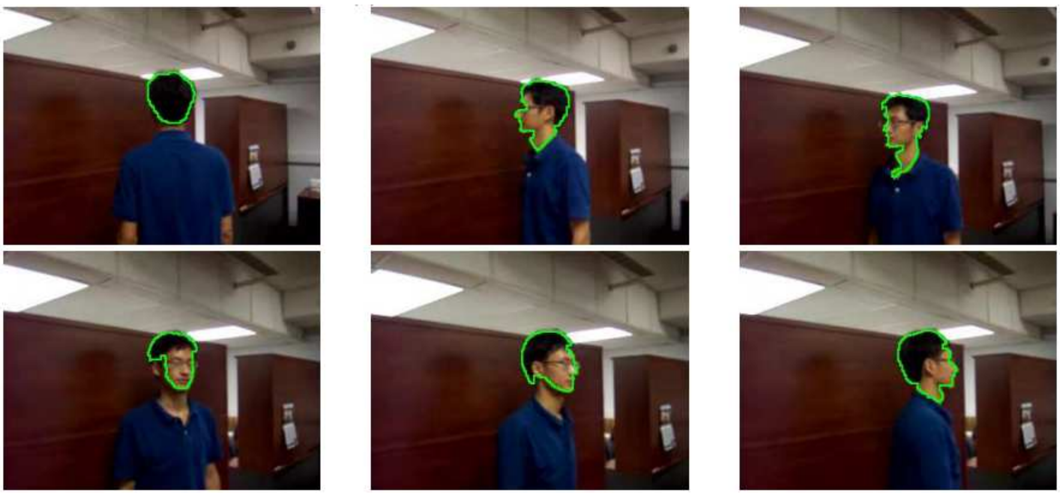


Figure 13: [1] The initial frame has only the hair color as part of the object model. When the person's skin color appears, the new fragment is added to the object model and hence the face of the person is also tracked

5 Work to be done

The work planned to be done next semester is to implement this adaptive fragments based tracking algorithm. It would be implemented on openCV platform and as discussed with sir, we are planning it to submit to openCV.

Evaluation process for the algorithm:

- A manual object contour location will be done, marking all the objects in a set of frames.
- We will run mean shift algorithm on those frames and find the ratio of overlapped region of object tracked by the algorithm with the actual object marked manually to the original region of the object as obtained by manual marking.
- Similarly we will obtain the same ratio using fragments based object tracking algorithm
- Compare the two ratios. The larger ratio signifies more accurate object tracking.

References

[1] Prakash Chockalingam. *Non-rigid multi modal object tracking using Gaussian mixture model.* PhD thesis, The Graduate School of Clemson University, 2009.

[2] Peter Meer Dorin Comaniciu, Visvanathan Ramesh. Kernel-based object tracking. *IEEE Trans. Pattern Analysis Machine Intell*, 25(5):564–575, 2003.

[3] Priwthwijit Guha. *Unsupervised Concept Aquisition From Surveillance Videos.* PhD thesis, Indian Institute of Technology Kanpur, 2009.

[4] Leow Wee Kheng. Mean shift tracking. Technical report, School of Computing, National University of Singapore.

[5] A. Mukerjee P. Sateesh, P. Guha. Colour and feature based multiple object tracking under heavy occlusions. *ICAPR*, 2007.

[6] Nalin Pradeep Prakash Chockalingam. Adaptive fragments based tracking of non-rigid objects using level sets. *International Conference on Computer Vision (ICCV)*, 2009.

[7] D.Ramesh V.Meer Venkatesh Babu, Comaniciu. Mean shift object tracking. Technical report, Indian Institute of Science, Bangalore, 2007.