Hybrid Tracking in OpenCV

Pushkar Kolhe

August 25, 2011

Abstract

1 Introduction

The mean shift algorithm was designed so as to search for a local probability density function (PDF) that approximates the empirical PDF. Mean shift has shown promising performance in many circumstances against image occlusions and clutters. In spite of its success in specific applications, however, mean shift has presented less efficiency in the presence of dramatic intensity or color changes in a pre-defined neighborhood. In these situations, additional features may be used as a complement that can improve the capability of the trackers. In our case we use a spare feature based tracker, such as SIFT or SURF. They can be integrated with the Mean Shift tracker for robust tracking.

2 Object description

2.1 For MeanShift

An object is described as a PDF. Object tracking involves searching for the most similar PDF across two neighbouring frames. In MeanShift the similarity measure is based on color features. We convert each frame into the HSV color space and use these channels for creating a PDF. For this hybrid tracking our features are in joint feature-spatial space.

$$p_x(x,u) = \frac{1}{N}\sum_{i=1}^{N} w\left(\left|\frac{x-x_i}{\sigma}\right|^2\right) k\left(\left|\frac{u-u_i}{h}\right|^2\right) \quad . \tag{1}$$

Here x is the position of a feature, u is the color feature. σ and h are used to normalize and w and k are kernels, in this case L2 Norms.

The MeanShift algorithm used for tracking works in the following way:

1. Create a histogram based on the chosen region of interest.
2. Backproject the histogram on the image.
3. Use the backprojection for Meanshift alongwith a distance measure.

2.2 Convergence of MeanShift

Meanshift is an instance of gradient ascent with adaptive step size. For asymptotic distributions close to the normal and gaussian distributions, convergence in meanshift is guaranteed. However color based PDF might not always have these asymptotic properties. MeanShift can get stuck at a local minima along the PDF. Integrating SIFT correspondence helps to make the object PDF asymptotic to the normal distribution, thereby reliably performing meanshift convergence at every frame.

2.3 For SIFT feature tracking

SIFT feature tracking works when the region of interest being tracked is textured or has many features. Contrary to MeanShift, which can track blobs of colors very well, SIFT feature tracking can track textured surfaces. Integrating them makes our tracker more robust.

$$p_x(x,u) = \frac{1}{N}\sum_{i=1}^{N} w\left(\left|\frac{x-x_i}{\sigma}\right|^2\right) f_s(x,u) \ , \tag{2}$$

$$f_s(x,u) = \frac{1}{2\pi\sigma_s^2} e^{-(d(x_{ms},x_{ft}))^2/2\sigma_s^2} \ . \tag{3}$$

Here $f_s(x,u)$ is the Gaussian kernel where the variance is the distance between the detected region centers using mean shifts and feature tracking.

The SIFT tracking algorithm works in the following way:

1. Find SIFT features on the region of interest.
2. Find SIFT features in the next image around the region of interest.
3. Find the best correspondence using SSD/RANSAC. Find displacement and scale of the new region of interest.
4. A gaussian PDF is created with the new tracked region's center as mean and the difference between MeanShift's predicted region and SIFT's trackers region as variance.
5. Use this Gaussian PDF alongwith distance measure as the object PDF.

2.4 Similarity search for the object PDF

A similarity search has to be performed on the object PDF in consecutive frames. This is achieved by using an EM algorithm that minimizes the difference between the mean shifts from the previous frame and the current frame. The object PDF consists of the two PDF's from MeanShift and SIFT matching. They are added in a linear fashion with weights given to each tracker. If the region of interest is textures, SIFT matching gets more weight. If the region of interest can be segmented as a color blob, MeanShift gets more weight. The weight are estimated by using linear regression.

$$p_x(x,u) = \frac{1}{N}\sum_{i=1}^{N} w_1 \left(\left| \frac{x - x_i}{\sigma} \right|^2 \right) k \left(\left| \frac{u - u_i}{h} \right|^2 \right) + w_2 \left(\left| \frac{x - x_i}{\sigma} \right|^2 \right) f_s(x,u) \ . \tag{4}$$

Here w_1 and w_2 are the two weights which are found using linear regression. They have the property that

$$\frac{1}{N}\sum_{i=1}^{N} w_1 \left(\left| \frac{x - x_i}{\sigma} \right|^2 \right) + w_2 \left(\left| \frac{x - x_i}{\sigma} \right|^2 \right) = 1 \ . \tag{5}$$

The overall algorithm for this hybrid tracker is as follows:

1. Define a region of interest in the first frame.
2. Compute the color histogram and backproject it on the image.
3. Generate a PDF based on color histogram and distance measure.
4. Find SIFT features and find best correspondence around the region of interest.
5. Generate a PDF based on the newly detected region of interest and the difference between the detected region using mean shift and SIFT matching.
6. Launch the EM algorithm and iterate until the difference between two meanshift is less than some threshold.
7. Launch linear regression to find weights of the two individual trackers.

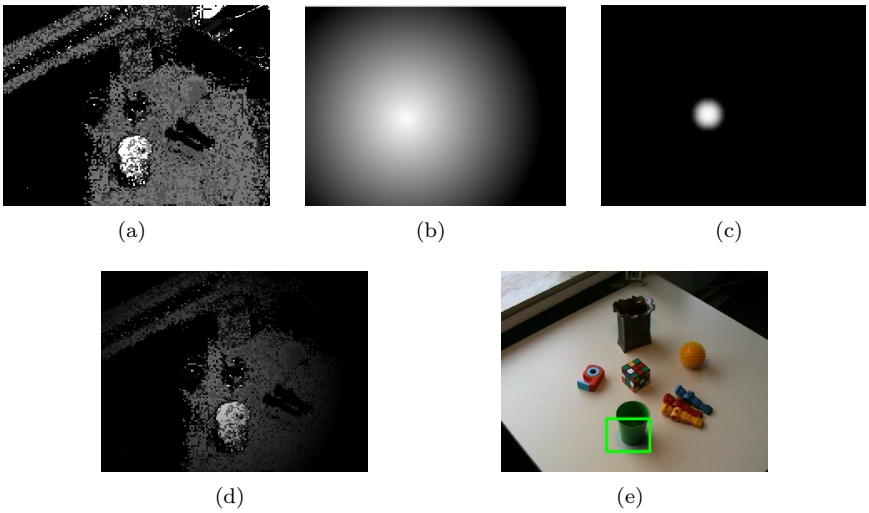
3 Experiments

The object PDF is generated by combining PDF's from individual trackers and adding the distance measure. The figure 1 shows the space in which each tracker visualizes the object. It is easy to notice that the MeanShift tracker relies heavily on color blobs, while the

4 Issues with the above approach

The above approach works well in clustered environments even in presence of occlusions. However, it does not track fast moving objects because it uses the EM algorithm to combine trackers which does not model motion. This can be improved by using a bigger window to search for SIFT correspondence, but EM will cause the hybrid tracker to not converge immediately.

To allow for other approaches, a motion model is added. The motion model can track some region of interest by using a Low Pass filter or the Kalman filter. Few examples are shown below.



(a) (b) (c)

(d) (e)

Figure 1: Object PDF. (a)PDF due to color historgram for MeanShift. (b) PDF due to the distance measure. (c) PDF due to the difference in MeanShift and SIFT correspondence. (d) Combined PDF for the object. (e) Object being tracked.

5 Results

The following were delievered as a part of this project:

1. A hybrid tracker based on Zhou et al.
2. An API/framework for tracking using MeanShift and Feature correspondence in OpenCV.
3. Sample code to use the trackers.
4. Sample code to benchmark the trackers.
5. Documentation for the framework and samples in reST docs.

6 Future Work

1. Add Ensemble tracking in this framework.
2. Use existing motion tracking algorithms based on Optical flow with OpenCV in this framework.