Paper Reading Summary

Raingo (raingomm@gmail.com)

March 29, 2011

1 Introduction

After days of paper reading, I have formed some basic understanding of some key algorithms which should be implemented in the project. Some of the basic algorithms actually have been implemented in the current OpenCV release.

2 MeanShift Algorithm

MeanShift, introduced in [2], is a basic algorithm that itetatively shifts a data point to the average of points in its neighborhood, with abundance application in cluster [1] and tracking. There are two implementation of this algorithm in current OpenCV release, for object tracking and for segmetation, respectively. Intuitively, the object of the meanshift algorithm is to find the mode nearest a specific point, as is shown in the Fig. 1 which is cited from [2].

In the tracking problem, the object function of meanshift algorithm is defined as the similarity measurement between the object model and current frame, and the object model is usually defined as the color distribution, which is denoted by the color histogram, as is suggested by the current OpenCV release that the probability matrix should be the backprojection of the frame.

Basically, based on the different assumption of the covariance matrix, there are two different version of meanshift, the one assuming the isotropic covariance, which is used in the [4] and current OpenCV release, and the one assuming the anisotropic covariance, which is used in the [3] and another algorithm named CamShift in current OpenCV release. Since the isotropic covariance assumption simiplify the object function of the meanshift algorithm, the anisotropic covariance version of meanshift turns out to more complex on the technical details with an addition itetative computation for the anisotropic convergence matrix. However, since the anisotropic convergence assumption takes the scale and orientation into consideration, it is more useful than the isotropic one.

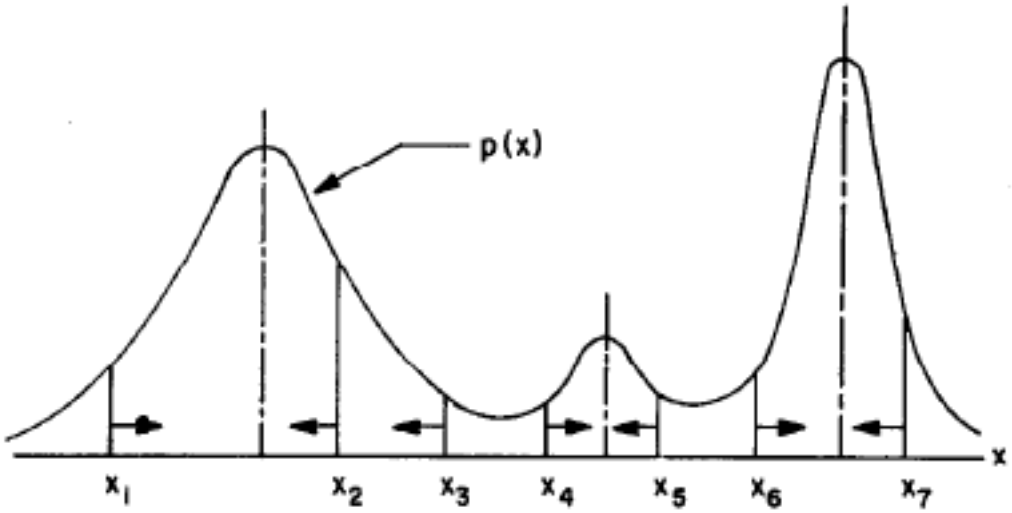

Fig. 1. Gradient mode clustering.

Figure 1: Figure cited from [2]

3 SIFT Algorithm

Intuitively, after several steps of image processing, the sift algorithm trys to extract the scale-invarient features [**?**], which has been implemented in current OpenCV release. These are features are used as the model of a specific target. Below is an outline of sift algorithm:

1. Constructing a scale space
2. LoG Approximation
3. Finding keypoints
4. Get rid of bad key points
5. Assigning an orientation to the keypoints
6. Generate SIFT features

After the SIFT features has been extracted, the descriptor can be used to identify the object with invarient in scale, rotation, illumination, viewpoint.

4 RANSAC Algorithm

Introduced in [**?**], RANSAC is widely used in robust parameter estimation. Intuitively, the fundamental thoughts of RANSAC is to extract underlie trend of the input samples,

which takes outliers into consideration, by means of randomly choosing small parts of inputs samples, not the whole samples to compute the parameters of the model.

After SIFT computation, there are always many matches between object model and current frame, however, the matches we want actually are not always the best matches with least feature distance between them, if so we can just sort the matches to extract the best matches we want. Since the SIFT is vulnerable about the noise, there are actually many outliers with good matches results, we can utilize the RANSAC to estimate the mostly likely location of object in current frame.

5 EM Algorithm

Introduced by [**?**], EM algorithm, which is utilized by the [4], is a technique used in point estimation.

EM algorithm itself is quite self-explanation, however, to tell the truth, I do not understand quite well its utilization in the paper [4] to maximize the similarity measurement concerning both meanshift classical measurement and the SIFT contribution.

6 Combination Scheme

It's easy to understand that combining the SIFT features, which have significent performance on object detection and meanshift algorithm, which utilize the characteristics of video flow, will significently improve the performance of object tracker. However, concerning how to combine these two algorithm, different people have different opinion.

In the paper [4], the combination appears in the object function of a optimazation problem. Maybe, we can refer to this scheme as a parallel scheme, which actuall merges the two different algorithm.

In the paper [3], the combination appears in the processure of object tracking, and we can refer to this scheme as a serial scheme, in which the object track problem has been divided into two relatively independant stages, utilizing SIFT algorithm to get a initial guess of the location of the object to start the meanshift algorithm on each frame.

In my opinion, both of the parallel scheme and the serial scheme have a perfect results on the object track task as is shown in the papers, however, both of them have huge computational cost and can not be utilized on the real-time scence, like robotics and real-time monitoring task. Based on the fact that if the target object has a good state during the tracking, which free from occlusion, intersection, deformation and pose changes, meanshift has a good performance on both speed and tracking results, which actually is the main state thourought the processing, we can combine the slow SIFT and the nice meanshift algorithm with another scheme. We compute the SIFT features only when the meanshift tracking results comes to a bad condition.

7 Conclusion

To conclude, after days of paper reading and days of writing this report, the framework of the joint tracking algorithm has been established in my mind, and I am excited to do the reseach things about the novel tracking algorithm, which actually provides a new perspective to my work.

References

[1] Y. Cheng. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8):790–799, 1995.

[2] K. Fukunaga and L. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Information Theory, IEEE Transactions on*, 21(1):32–40, 1975.

[3] Z.H. Khan, I.Y.H. Gu, T. Wang, and A. Backhouse. Joint anisotropic mean shift and consensus point feature correspondences for object tracking in video. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 1270–1273. IEEE, 2009.

[4] H. Zhou, Y. Yuan, and C. Shi. Object tracking using SIFT features and mean shift. *Computer Vision and Image Understanding*, 113(3):345–352, 2009.