Hand Tracking Using Distinct OpenCV Methods and Analysis of their Accuracy

Francisco Jose Nardi Filho
Lucas Gasparetto Farris

June 24, 2015

Abstract

Mean Shift, Camshift and KLT are three successful approaches to visual tracking. However, each of them have their particular strengths and weaknesses. In this paper, we intend to present implementations of each distinct method from the OpenCV library and make an accuracy and performance analysis for a hand tracking case. Fifteen videos taken from YouTube showing diverse situations as teaching of the American signal language alphabet, Mudras (gestures from an Indian dance), as well as guitar playing were used as dataset to run the algorithms and perform the hand tracking. The hand position on these one-minute videos was manually annotated prior to the methods' accuracy and performance analysis. Extensive experimental results demonstrate that XXXXXXXXXX outperforms YYYYYYYYYY in hand tracking considering the provided database. Our approach produces reliable tracking while effectively handling rapid motion.

1 Introduction

Object tracking is to determine the position of the object at each frame of a video sequence. Reliable visual tracking is indispensable in several fields such as: ambient intelligent systems, augmented reality, human–computer interaction, video compression, and robotics. However, the task of robust tracking is very challenging regarding fast motion, occlusion, structural deformation, illumination variation, background clutters, and real-time restriction. Tracking algorithms can be divided into two categories.

The first category is deterministic method. Mean-Shift is a typical example. This method finds the local maximum of probability distribution in the direction of gradient. It is usually more accurate and quick in tracking than the probabilistic multi-hypothesis tracking algorithm. But it may run into trouble when similar objects are presented in background or when complete occlusion occurs. Camshift is called the Continuously Adaptive Mean Shift algorithm which is a modified algorithm of Mean-Shift.

The second category is probabilistic method. The representative method is particle filter which is a multi-hypothesis tracking algorithm under the Bayesian framework . Due to particle filter's non-Gaussian non-linear assumption and multiple hypothesis property, they have been successfully applied to visual tracking, and show unique merit in cluttered environment. However, the inefficiency

in sampling (due to the problem of degeneracy and impoverishment) and the huge computational complexity limit the usefulness of particle filter in on-line tracking [6].

Therefore, each method has its own advantages and disadvantages, manners to overcome the challenges described before. Aiming to find the optimal OpenCV method for hand tracking, and being aware of the nature of their categories, we have chosen four techniques very present in the literature of the state-of-art hand tracking, which are: Mean Shift, Camshift using SIFT keys, KLT and Particle Filtering.

Videos extracted from YouTube having about 1 minute of duration and showing diverse situations as teaching of the American signal language alphabet, Mudras (gestures from an Indian dance), as well as guitar playing were used as dataset to run the algorithms and perform the hand tracking.

A metric based on the matching between the annotated hand position on the videos and the hand position that was identified by the algorithm was defined in order to analyze the the algorithm's accuracy. Regarding the performance, the analysis took into consideration the time and space used by the algorithms.

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Extensive experimental results demonstrate that XXXXXXXXXXXX outperforms YYYYYYYYYYYY in hand tracking considering the provided database. Our approach produces reliable tracking while effectively handling rapid motion.

2   Related work

There have been done many works in hand tracking. [1] designed a fast tracking algorithm that combined Kanade-Lucas-Tomasi(KLT) flocks and k-nearest neighborhood. Based on the pros and cons of particle filter and mean shift, [5] proposed a new algorithm, Mean Shift Embedded Particle Filter (MSEPF), in which mean shift is performed on each of the particles after they are propagated, so that the particles are "herded" to nearby local modes with large probability. Thus, the problem of degeneracy, where the weights of most particles become negligibly small after a few iterations, is tackled effectively. It is also claimed that as MSEPF can make better estimation of posterior even with a smaller set of samples, the computation cost is reduced proportionally. [2] also proposes a hybrid particle filtering and mean shift tracking algorithm, however using one more feature: an adaptive state transition model with updating variances. The update process is driven by the variability of the target in previous frames.

In another perspective, [4] presents a scale and orientation adaptive mean shift tracking (SOAMST). It addresses the problem of how to estimate the scale and orientation changes of the target under the mean shift tracking framework. In the original mean shift tracking algorithm, the position of the target can be well estimated, while the scale and orientation changes can not be adaptively estimated. Considering that the weight image derived from the target model and the candidate model can represent the possibility that a pixel belongs to the target, it is shown that the original mean shift tracking algorithm can be derived using the zero th and the first order moments of the weight image.

There are still some works as in [3], which performs hand detection first extracting the images of hands (or palms) by mapping a circle plate in the candidate regions of hand and the center point of circle of same skin color. The

mapping is done by employing Bresenham's Midpoint circle scan-conversion algorithm for identification of the regions that cover the palm. As soon as the area matches the circle, it is added an initial position condition relative to the face to filter other noise regions before starting hand tracking. The hand detection method is applied on the fixed search window region for hand tracking. The boundary of hand is calculated and the centroid point of hand region is determined. Through iteration of hand tracking process, it is possible to obtain the motion trajectory of the hand so-called gesture path from connecting hand centroid points set.

3   Experiments

In this section, we carry out experiments to evaluate the accuracy of the developed hand tracking algorithms on 15 video sequences. The majority of the videos show a single hand most part of the time. However, there are many changes in the camera focus, which increases difficulty for the algorithms to keep track of the hand correctly.

Once the videos were annotated, the center of the hand in every frame, in every video is known. Therefore, the hand itself is represented by a circumference of fixed radius. In turn, the algorithms return a convex polygon with the region they understand as being the hand's most probable location. A grade 1 or 0 is assigned to the algorithm's score, 1 if the polygon's region matches the circumference's, 0 if it does not. AFTER ....

To investigate how much the Camshift/ Mean Shift .... iteration can improve the sampling efficiency of PF for hand tracking, we also performed comparative study on the number of particles between .

resents the tracking results when varying the number of particles. It is observed that, as hand motion varies much and a weak dynamic model is adopted, PF requires at least 150 particles for stable tracking in all test sequences, while MSEPF requires only 20 particles. 1 With re- gard to the computation time, PF (150 particles) spends 63 ms on average for each frame, whereas for MSEPF (20 particles), though MS iterations added, each frame only costs an average time of 28 ms as the number of particles is decreased greatly.

Therefore, in the application of hand tracking, MSEPF de- crease 85Table 2 summaries the optimized tracking performance of the three algorithms on all test sequences. We observe that MSEPF achieves better tracking results than both PF and MS. Although it is faster than MSEPF, MS cannot achieve stable tracking and sometimes loses the track.

4   Results and discussion

5   Conclusion

References

[1] M. Kolsch and M. Turk. Fast 2d hand tracking with flocks of features and multi-cue integration. In *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 10 -*

Volume 10, CVPRW '04, pages 158–, Washington, DC, USA, 2004. IEEE Computer Society.

[2] E. Maggio and A. Cavallaro. Hybrid particle filter and mean shift tracker with adaptive transition model. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 2, pages 221–224, March 2005.

[3] B. Min, H. Yoon, J. Soh, Y. Yang, and T. Ejima. Hand gesture recognition using hidden markov models. In *Systems, Man, and Cybernetics, 1997. Computational Cybernetics and Simulation., 1997 IEEE International Conference on*, volume 5, pages 4232–4235 vol.5, Oct 1997.

[4] J. Ning, L. Zhang, D. Zhang, and C. Wu. Scale and orientation adaptive mean shift tracking. *Computer Vision, IET*, 6(1):52–61, January 2012.

[5] C. Shan, T. Tan, and Y. Wei. Real-time hand tracking using a mean shift embedded particle filter. *Pattern Recognition*, 40(7):1958 – 1970, 2007.

[6] M. Yin, J. Zhang, H. Sun, and W. Gu. Multi-cue-based camshift guided particle filter tracking. *Expert Systems with Applications*, 38(5):6313 – 6318, 2011.