INTERNATIONAL JOURNAL OF RESEARCH IN COMPUTER APPLICATIONS AND ROBOTICS
ISSN 2320-7345

EFFICIENT HUMAN TRACKING SYSTEM

S.R.Bharani Alankar[1], C.S.Manikanda Babu[2]

[1]PG Scholar, Sri Ramakrishna Engineering College-Coimbatore,
bharanialankar@gmail.com
[2] Asst.Professor(Sr.Gr), Sri Ramakrishna Engineering College-Coimbatore
srecmanikandababu@gmail.com

Abstract

This paper addresses the key concerns to track several subjects from video sequences acquired by multiple cameras in real time and continuity of tracking in overlapping and non-overlapping fields of view. Each human subject is referred by a parametric ellipsoid having a state vector that encodes its position, velocity and height. We also encode visibility and persistence to tackle problems of distraction and short-period occlusion. To improve likelihood computation from different viewpoints, including the relocation of subjects after network blind spots, the colored and textured surface of each ellipsoid is learned progressively as the subject moves through the scene. This is combined with the information about subject position and velocity to perform camera handoff. For real time performance, the boundary of the ellipsoid can be projected several hundred times per frame for comparison with the observation image..

Keywords: Multiple Object Tracking Accuracy (MOTA)

1. Introduction

The project tracks subject in specified region from video sequences acquired by camera in real time. Each human subject is represented by a parametric ellipsoid having a state vector that encodes its position, velocity and height. We also encode visibility and persistence to tackle problems of distraction and short-period occlusion. To improve likelihood computation from different viewpoints, including the relocation of subjects after network blind spots, the colored and textured surface of each ellipsoid is learned progressively as the subject moves through the scene. For real time performance, the boundary of the ellipsoid can be projected several hundred times per frame for comparison with the observation image. The system includes the observation and tracking of human subjects as they move within the field of view of camera networks. For example, CCTV surveillance is used to record and counteract criminal acts in town centres, public buildings and transport terminal, for traffic monitoring to apply congestion charging, for road planning, and to observe shopping patterns in a supermarket. In the context of CCTV, we aim to deploy economically distributed sensing and computation with cooperative communication, for example to track an identified individual from camera to camera in an urban setting. Real time processing is a priority if responsive action is required before rather than after an event.

2. Background and related work

A literature survey is done for various papers which are essential to know the previously available techniques and their significance and limitations. It also includes the various supporting papers for the proposed technique and their advantages. They consider as Multiple Human Tracking Accuracy (MOTA) for speed of processing in tracking system.

D. Comaniciu and P.Meer Proposed "Mean shift: A robust approach toward feature space analysis,". The basic idea is that the output is state vector is estimated by finding the optimal state which maximizes a likelihood function which includes the appearance descriptor.

A. Jepson, D. Fleet, and T. El-Maraghi Proposed "Robust online appearance models for visual tracking". This paper proposes a robust, adaptive appearance model for motion-based tracking of complex natural objects. The model adapts to slowly changing appearance, and it maintains a natural measure of the stability of the observed image structure during tracking.

Z. Kalal, K. Mikolajczyk, and J. Matas Proposed "Face-TLD: Tracking learning-detection applied to faces". There is the additional problem of data association, both within and between camera views, linking detected subjects with associated state vectors to trajectories. The proposed also Detections in space and time can be linked to shortest path by edges and boundaries.

M. P. S. N. Sinha, J.-M. Frahm, and Y. Genc Proposed Feature tracking and matching in video using programmable graphics hardware. To implemented a parallel feature detector and tracker on a GPU achieving speedups of approximately 10 compared to an optimized CPU implementation. This is comparable to the performance improvement we expect in our work, but we address a rather different problem of detection and tracking of extended bodies

K. Pauwels, M. Tomasi, J. Diaz, E. Ros, and M. V. Hulle Proposed "A comparison of FPGA and GPU for real-time phase-based optical flow, stereo, and local image features". The development of GPU technology lags behind that of FPGAs in their restricted capability for deployment in safety critical areas, and they have relatively high power consumption. They consisted of advantages of development time and higher levels of abstraction.

R. Z. M.Wojcikowski and B. Pankiewicz, Proposed the FPGA-based real-time implementation of detection algorithm for automatic traffic surveillance sensor network. To implemented an FPGA detection algorithm for traffic surveillance as part of a sensor network. Vehicle motion is much more regular, and the consequent solution is implemented as repeated detection using static data structures with a high degree of pipelined data parallelism.

G. B. J. Sanchez and J. Simo in Video sensor architecture for surveillance Applications. To used a hybrid architecture of DSP and FPGA devices for video surveillance. A maximum response that is also above a heuristic threshold of 0.65 is a new subject.

From the above literature work, it is evident that the previous techniques have used state vector and texture details separately for detection and tracking. Hence the new algorithm is needed to detect subjects with 3-D ellipsoid and texture signature combined to estimate state variables. Also the processing speed and accuracy should be improved in order to achieve real time requirements..

Tracking should be includes Prediction and evaluation. In Prediction includes a motion model of space in succeeding video frames. Evaluation requires a comparison between the prediction of position, velocity and appearance and the observed data. The existing work presents different State estimators; maximum likelihood (ML), maximum a posteriori (MaP), Bayesian estimation.

2.1 Nadaraya-Watson Estimator

The location estimator is model of density estimated with the Kernel K from the available data. Mean shift procedure to move toward the mode (peak) to analyze such spaces. Kernel density estimation technique is most popular density method. This estimator is measured by the mean of the square error between the density and its integrated. Non-parametric method to estimate complex from noisy data.

2.2 Robust M-Estimator

The M-estimators are family of robust techniques which can handle data in the presence of serve contaminations. The location estimator is mode of density with the Kernel K available data. The relation between location of M-estimators and Kernel density estimation is not statistical literature. It context of an edge preserving smoothing technique. The classification error similar to optimal Bayesian classifier. The pruning of the mode candidates produced seven peaks.

2.3 EM - Estimator

To estimate the parameters of the generative model in (a), namely, the mean and variance of the data prediction by the stable process, q . ._s; _2s ., and the mixing probabilities m . .mw;ms;ml.. Since we plan to apply this mixture model estimation scheme to high dimensional appearance data, like the responses of a wavelet filter bank, it is very important that we find an efficient computational algorithm that requires a small amount of memory for each temporal stream of data observations.

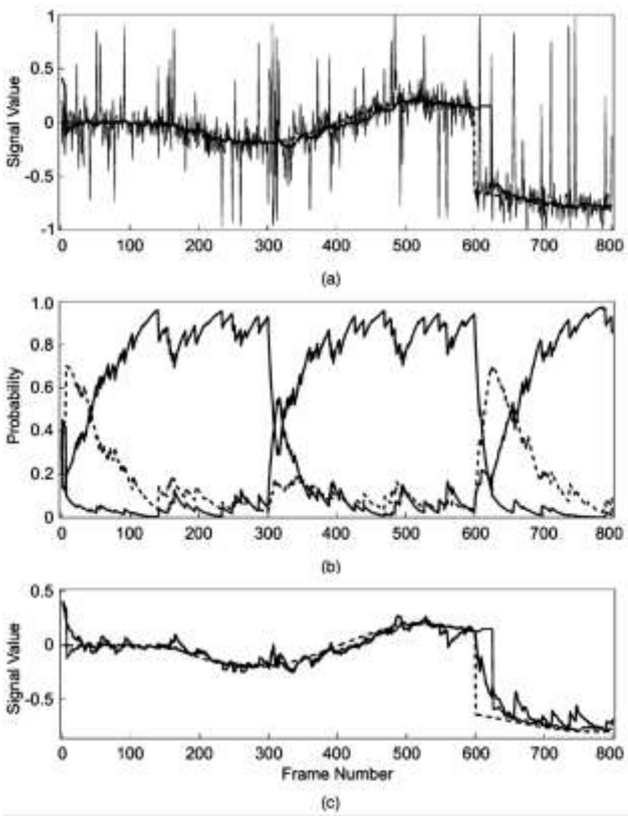

Figure 1 EM estimator a) input image, b) mixing probabilities, c) IIR filter output

Finally, by way of comparison, conventional adaptive templates typically use a recursive (IIR) linear filter to average the data observations. In this way, the WSL mixture model achieves our goal of capturing the structure of slowly varying signals having both outliers and occasional sudden changes in appearance. The WSL mixture model that combines predictive density models of appearance with components that adapt over long and short time courses, an online version of EM for learning the parameters of the WSL model, an application in which we learn the time-varying phase behaviour of a steerable pyramid, and a tracking algorithm which exploits this appearance model to simultaneously estimate both motion and appearance.

The above discussed estimator provides tracking based on present or future constraints separately . Every estimator has their own advantages and limitations. To design an accurate and efficient estimator is important in the case of real time applications. Finally, Bayesian estimation is used to produce better result

3. Proposed Technique

In this paper, we describe a Bayesian approach to track multiple human subjects within a camera network that can have both overlapping and non-overlapping fields of view. The contributions are as follows. First, we introduce a new form of appearance model, based on a colored, textured 3-D ellipsoid that is progressively learned as the subject moves through the network.

3.1 Tracking in an overlapping Network

The main idea of overlapping is track a initiation of a subject to track performed independently in each camera view. The prior density of position is assumed to be a uniform distribution within a 1 m radius on the ground plane. The SIR-PF has transition, likelihood and re-sampling functions. The *likelihood* is based on the similarity between a synthetic appearance generated from a particle state using the ellipsoid model and the observation image. *Resampling* draws samples based on *importance sampling* and is implemented by inversion sampling.

3.1.1 Silhouette likelihood

The ellipsoid is a closed surface in 3-D space. Any surface coordinate is mapped to a cylindrical coordinate system, which allows a 2-D image of an ellipse boundary in each camera image defined by the cylinder azimuth and elevation and camera calibration. The rays from all active foreground pixels are traced and the intersection with the ellipsoid surface is computed quickly. The number of intersections as a fraction of the ellipse area forms the silhouette likelihood.

3.1.2 Texture likelihood:

The texture distribution is an accumulated average over time acquired by mapping from the image to the ellipsoid surface coordinates. The facing angle of the subject is assumed in the direction of velocity to define the azimuth

3.1.3 Combined likelihood for a single camera:

The combined likelihood is a linear combination of the silhouette and texture likelihoods, incorporating distraction suppression

3.1.4 Combined likelihood for a multiple camera:

The likelihoods from individual cameras are evaluated for reliability before forming a combined likelihood

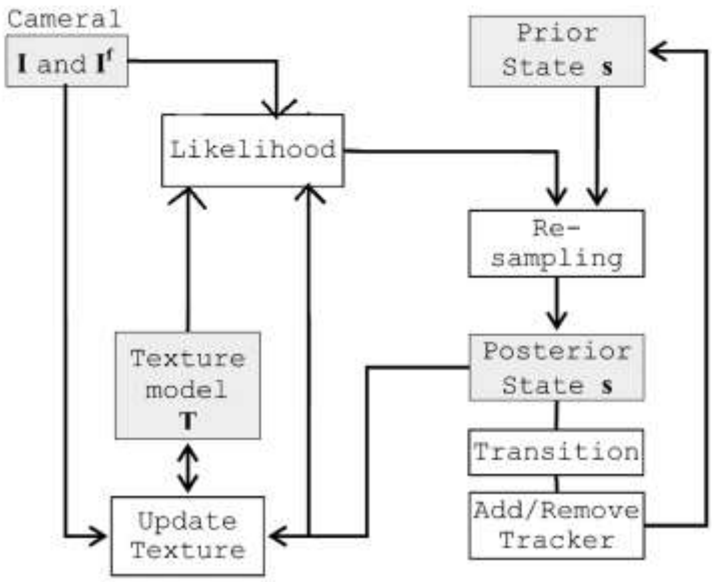

Figure 2: Schematic of Tracking Module

3.1.4 Feature Comparison:

Once the required feature (pointer position)is extracted from the image then the respective angle value is assigned to the pointer. the present value of the pointer is compare with the set value. If the obtained pointer's value exceeds the limit then alarm will be turn on.

3.2 Evaluation of overlapping network:

We have measured the accuracy, speed of computation and execution profile. The sequential implementation is coded in C++ and tested on a single core of a CPU Intel Core i7 950, clock frequency 3.1 GHz. We have used several public datasets including PETS03 and PETS06, PETS09 and two of our own datasets, EM330 and DEC11. Three frames from two views in the PETS09 data set. To set the several weighting factors in (10), we used the PETS09 data as a training set, choosing the optimal set to obtain the highest MOTA, but then used these parameters.

We compare these results with those reported for the same dataset. The MOTA of our tracking algorithm is higher than the best method. First, the 3-D ellipsoid model allows observations from multiple cameras to be integrated effectively.

Second, our method uses detection to activate the PF and then the 3-D ellipsoid with an evolving texture signature to estimate the state variables. Thus, the state variable and texture signature are combined into a subject representation for likelihood calculation, unlike for example, who considered state variables and appearance separately. Other authors have also linked detected subjects to trajectories based on 2-D features, which is difficult to extend from a single to multiple cameras. Here, using the ellipsoidal texture learning method, the signatures are defined consistently across all cameras, resulting in the higher MOTA values.

3.3 Evaluation of non-overlapping network:

We use the *F*-measure that combines the *recall* and *precision* rates defined. This is computed from the number of correct assignments estimated by our algorithm and the total number of assignments of ground-truth. We use both the public and our private datasets *with* camera calibration. Combining spatial and texture data association, there are three tuning parameters, the standard deviation of the motion model, the threshold, and

Vol.2 Issue.3, Pg.: 21-28
March 2014

blending factor. The effectiveness of camera handoff is critically dependent on the uniqueness of the attire (which affects the discriminative ability of texture comparison) and the regularity of movement (which affects spatial association).

3.4 Tracking in a Disjoint network:

Now address the camera handoff problem, tracking and identifying a subject when he/she has disappeared from sight to reappear in at least one camera view. When that subject reappears, this is registered as detection, and so the subject must be identified either as a new or a previously viewed subject. We use spatial and texture information to possibly link the broken trajectory and assign an existing or new ID.

Tracking should be includes Prediction and evaluation. In Prediction includes a motion model of space in succeeding video frames. Evaluation requires a comparison between the prediction of position, velocity and appearance and the observed data. The existing work presents different State estimators; maximum likelihood (ML), maximum a posteriori (MaP), Bayesian estimation.

The comparison of proposed system in overlapping and non-overlapping fields with detection and likelihood feature of much lesser extent is achieved. Also it gives better speedup and contributes to the overall reduction in computation time. Finally we observed that the processing time for implementing new algorithm is very less

4. Implementation and Result

The output video tracks the human in the specified area. The state vector with ellipsoid red color marker shows the tracking of person regardless of the direction. also the tracking performed for the specified portion of the video. In this video only the centre part is considered for tracking.

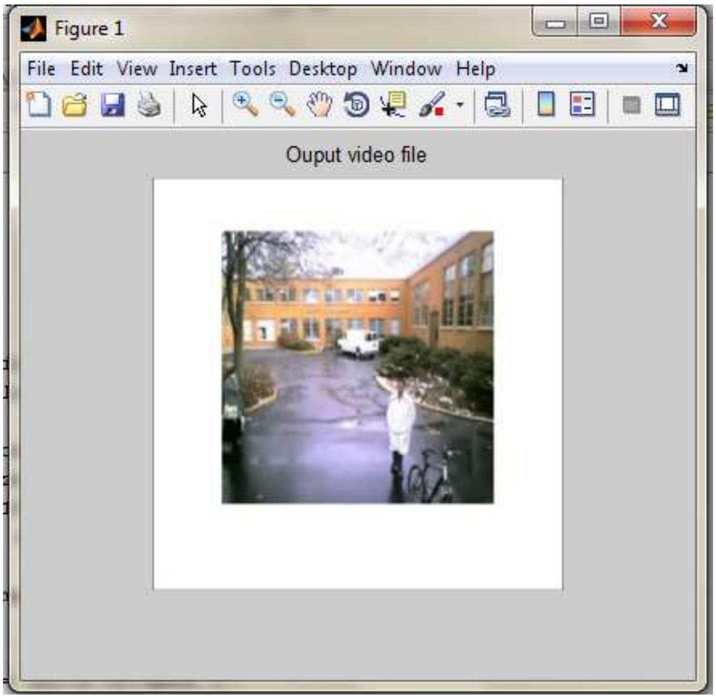

Figure. 3. Output Video

5. Conclusion

The proposed approach is to track human subjects in specified regions of video sequences . First, we have introduced a parametric ellipsoid model in both detection and visual tracking. For detection, this is projected at static grid positions to find intersections between potential subject positions and foreground image data, as determined by mixture of Gaussian segmentation. For tracking, the ellipsoid is parameterized by position, velocity and height as part of the state vectors of a particle filter.

Table 1 Comparison of Estimators

Estimator	Constraint Dependence	Accuracy
ML estimator	Present	Low
MaP estimator	Present and future	Medium
Bayesian estimator	Present ,future and predicted values	High

The following Table 1 shows the advantage of Bayesian estimator over the existing estimators. the present estimator i.e., Bayesian estimator provides better results by combining advantages of two other estimators. Bayesian estimator combines both advantages of ML and MaP also includes predicted values. These predicted values compares with previous results for better results. In future, the implementation of the project can done using GPU in order to achieve real time performance tracking of many subjects. The speedup is achieved using benchmarks.

REFERENCES

[1] D. Comaniciu and P. Meer, May 2002, “Mean shift: A robust approach toward feature space analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 5, pp. 603–619.

[2] A. Jepson, D. Fleet, and T. El-Maraghi, Oct. 2003, “Robust online appearance models for visual tracking,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 10, pp. 1296–1311.

[3] [9] Z. Kalal, K. Mikolajczyk, and J. Matas, Sep. 2010, “Face-TLD: Tracking- learning-detection applied to faces,” in Proc. 17th IEEE Int. Conf. Image Process., pp. 3789–3792.

[4] M. P. S. N. Sinha, J.-M. Frahm, and Y. Genc,2011, “Feature tracking and matching in video using programmable graphics hardware,”Mach.Vis. Appl., vol. 22, pp. 1207–217.

[5] K. Pauwels, M. Tomasi, J. Diaz, E. Ros, and M. V. Hulle, Jul. 2012 ,“A compar- ison of FPGA and GPU for real-time phase-based optical flow, stereo, and local image features,” IEEE Trans. Computers, vol. 61,no.7, pp.999–1011.

[6] R. Z. M. Wojcikowski and B. Pankiewicz, Jan. 2012, “FPGA-based real-time implementation of detection algorithm for automatic traffic surveillance sensor network,” J. Signal Process. Syst., vol. 68, no.1, pp.1–18,.

[7] G. B. J. Sanchez and J. Simo, May 2006 , “Video sensor architecture for surveillance applications,” Sensors, vol. 12, pp. 1509–1528, 2012 [8] G.G.Acosta, C.J.Verucchi, and E.R.Gelso,“A current monitoring system for diagnosing electrical failures in induction motors,”Int.J. Mech.Syst. Signal Process, vol. 20, no. 4, pp. 953–965.

A Brief Author Biography

S.R.Bharani Alankar – PG Scholar at Sri Ramakrishna Engineering College, Coimbatore, India. Research interest is Image processing.

C.S.ManikandaBabu – Assistant Professor (Sr.Gr) at Ramakrishna Engineering College, Coimbatore, India. Research interest is Image processing.