International Journal of Computer Science and Engineering (IJCSE)
ISSN 2278-9960
Vol. 2, Issue 3, July 2013, 107-114
© IASET

International Academy of Science, Engineering and Technology
Connecting Researchers; Nurturing Innovations

HUMAN TRACKING AND POSE ESTIMATION IN VIDEO SURVEILLANCE SYSTEM

NEELAM V. PURI[1] & P. R. DEVALE[2]

[1]PG Scholar, IT Department, Bharti Vidyapeeth College of Engineering, Pune, Maharashtra, India

[2]Professor & Head, Department of Information Technology, Bharati Vidyapeeth College of Engineering, Pune, Maharashtra, India

ABSTRACT

Video surveillances system based on human detection, tracking and human pose estimation assurances to be an important technology for real time applications, including the analysis of human activities. So many applications have been demonstrated regarding this technology but evaluations of some key features are remains challenging. How human being going to be detected that is the challenge i.e. According to body structure, skin color, skeleton, etc. It is very complex because each human being has different human kinematic structure, variation in body size and shape. In the human tracking process occlusions of body parts, inabilities to observe the skeletal motion due to clothing, difficulty segmenting the human from the background these are the challenges. Pose estimation is also challenging because complex interactions between people in the environment, clothing complicates the skeleton structure, and significantly increases the inconsistency of individual human appearance. Some image related components also increases the challenges because limited image resolution, number of ambiguities, and the inability to easily distinguish the parts of a human from occlusion or from the cluttered background. With the prior knowledge some of these challenges are resolved, but some of the problems require clever mathematical and engineering solutions.

KEYWORDS: Discriminative Methods for Pose Estimation, Human Detection, Human Tracking, Kalman Filter for Tracking, Pose Estimation, Histogram of Oriented Gradient Algorithm

INTRODUCTION

This paper concerned with the task of human tracking in video sequences. Given video stream from a stationary camera, it was desired to detection of foreground objects as human being, at different frames of a sequence and labeled with boundaries.

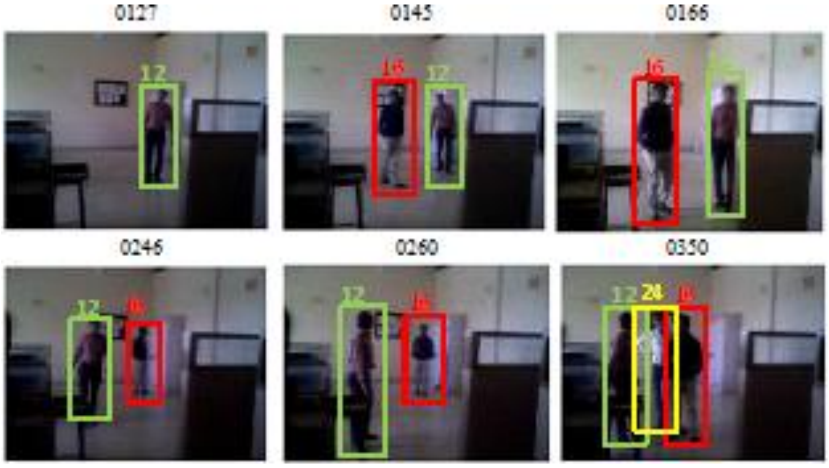

Figure 1: Tracking Objects through a Video Streams. Tracked Objects are Shown at Different Frames of a Sequence and Labeled with Boundaries

And track the same object moving through the scene as example in Figure 1. After algorithm works for pose estimation and activity analysis.

Previous Work Done

Various detection and classification techniques are available in this field of surveillance, with each system using its own technique. Ross Cutler and Larry S. Davis [1] developed a surveillance system attempts to detect and analyze the periodic motion using Time Frequency Analysis in a moving entity.

It stabs to detect and classify the object into humans, vehicles, animal etc.Liable on the periodic motion. For example, a running animal exhibits significant periodic motion in his legs regions. The System introduced by Ismail Haritaoglu, David Harwood and Larry S. Davis [2] is enormously effective system stated for detection and tracking of human in a video stream. Using a statistical background model it segments out the foreground pixels and describes them as blobs. Using static silhouette shape and dynamic periodicity analysis the blobs are categorized either in single person or people in a group or in other objects. Another System presented by N. Dalal and B. Triggs [3] works on static images for recognizing humans using Histograms of Oriented Gradients. The algorithm is based on evaluating normalized local histograms of image gradient orientation in a dense grid. The method tries to characterize local object appearance and human shape by a general idea of the distribution of local intensity gradients or edge directions. Q. Zhu, S. Avidan, M-C Yeh, and K-W Cheng [4] introduced algorithm for systematically search for an object in an image, used a Cascade of Histogram of Oriented Gradients to search for humans in an image. Particular this algorithm uses search windows at various scales and locations in the image. A video sequence captures around 15 to 25 images/second to search in, depending on the fps of the video.

Hence, adapting already existing object recognition algorithms is required for easily recognizing objects in static images by converting video into images.

Proposed Video Surveillance System

This paper represents detection of human being using the Histogram of Oriented Gradient Algorithm [4]. Human tracking is done based on work by Stauffer and Grimson [5]. Additionally, work towards an appearance-based object model based on Connor and Reid [6] is undertaken. Video sequence is converted in sequences of images to detect foreground object reference image is compared with current image and each pixel is labeled as either foreground or background. The next is to link object observations at different frame in a sequence to capture the object's motion path. This system is resilient against temporal illumination changes. The system also gets used to itself to long lasting changes in the background over time. Pose estimation stage primarily following monocular pose tracking with a probabilistic formulation with minimal occlusion. The problems faced in monocular tracking often arise in the general multiview case as well.

FOREGROUND DETECTION

In first stage of the system, live video stream is recorded with 15-25 frames per second rate. This video stream then converted into image sequences. Only condition for this system is stable capturing input device (i.e. Background of the video should be stable). Moving object then will be detected as foreground. Moving object detection computes the foreground/background pixel labeling for each sequence image. When previous image frame will be compared with current image frame, it will identify pixels that are sufficiently unusual in the context of their previous values. To find out difference in pixel values a distinct background model i.e. Mixture of isotropic Gaussians is used. Which captures the distribution of recent image pixel values is maintained for each pixel. As the sequence continues for next image frame the mixing proportions, means and variances of the Gaussians are updated through a set of rules that progressively push likelihood frame towards recently detected pixel values, thereby absorbing new changed pixel value. Change in recent pixel value is captured either based on current parameters of pixel a set of three recursive update equations which modify a single Gaussian's statistics, or by replacing a Gaussian with one centered at the current pixel.

When a new image comes for observation, the algorithm compares each pixel to its reference background model to decide foreground/background membership by a method approximating outlier detection. An example input image and the computed labeling is shown in Figure 2. Foreground objects are extracted from a binary labeling as large one connected components. This component is called blobs.

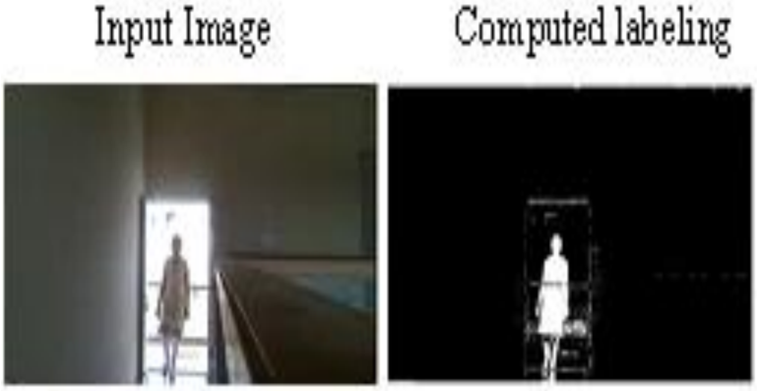

Figure 2: According to Reference Image, the Algorithm Detects the Binary Foreground/Background Labeling

To implement Stauffer and Grimson's [5] basic algorithm it’s necessary to understand some of the key properties of it, according its modification and analysis. Central dynamics is the combination of graphical views and recursive adaption of various statistics. Graphical views were developed that facilitated manual inspection.

When algorithm is running on new image sequence new hitches arises between algorithm, parameter, and sequence statistics. A number of modifications were proposed and evaluated during the decision making process of an algorithm.

For example Stauffer and Grimson's [5] proposed mixed component should be replaced entirely, but it was not clarified precisely which one. Total three replacement strategies are evaluated according to observing their parameters with the ground truth respect to a segmentation loss function. With the ground truth data the function captures to what degree the computed labeling agrees.

From random initial parameters several optimizations were carried out, and finally all three techniques compared. To find out flexible mixture component, diagonal instead of spherical Gaussians, combination of same three strategies are used with appropriate update equations. The optimization method that was developed evidenced beneficial for parameter changes of pixels values on different image sequences.

When working in interested domain on Gaussian mixture components for understanding of recursive equation updates system raises some new challenges.

Solution for this is the recursions were rollback and rewritten in closed-form. A component statistics are an exponentially weighted average of the mixing proportion and the mean Tor of the variance of the Gaussian.

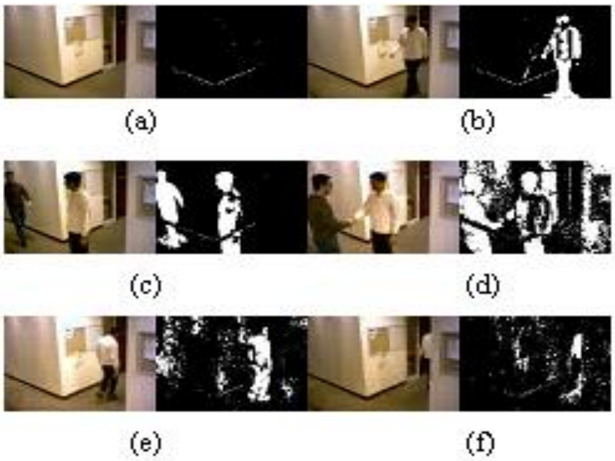

Figure 3: Experimental Frames Showing Various Computed Foreground Labeling during Day-Time. The System Adapts to Varying Lighting Levels and Changing Scene Geometry

That means average of initial estimation and the true process statistic. Authors suggested a rough estimation of a time constant in terms of parameters to the equations for describe the algorithm dynamics. Using a Taylor expansion for reasonable parameter ranges approximate time constant derived, and these derivations allowed determining actual time constants of key parameters that control the algorithm's dynamics. Algorithm works on the evolution of a single pixel's background model when it is being exposed to slow but sure illumination changes with foreground object movement.

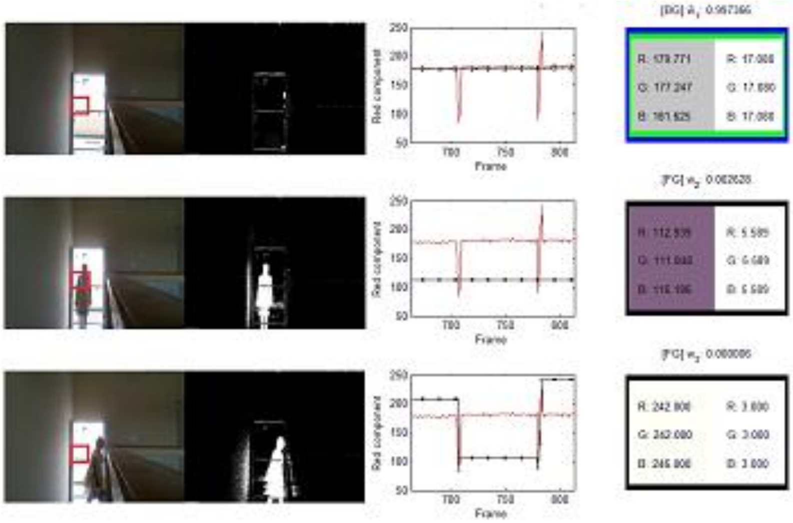

Figure 4: First Column Shows Three Example Image Frames of a Video Sequence, Second Column Shows the Human Detection as Foreground Object in Binary form, Third Column Shows the Evolution of Three Mixture Components during that Video Sequence, and Last Column Shows the Foreground/Background Detected Color Values [7]

The previous notions were made precise by a case study of the evolution of a single pixel's background model while being exposed to steady illumination changes and the passing of a foreground object.

While adequately detecting foreground objects as human being for checking the robustness against changes natural background scenes the algorithm was run on a live video stream for 1 hour and some of the results shown in Figure 3. This experiment confirms the Stauffer and Grimson's [5] claim. The inconsistency detected in RGB data of background scenes i.e. Change in lighting. In first column of Figure 4 three sequential frames converted from the video stream are shown. The interested area is selected for analysis purpose and highlighted with red outlined box. Second column shows binary frames with detected human movement. The evolution of the three mixture components for frames 633 through 813 are shown in third column. The earlier analyzed dynamics were made concrete for a sequence.

OBJECT TRACKING

[7] It is very hard to identify foreground object as human being in different time period because of the light shades and all natural changes. Foreground object can be detected as human being using computed labeling in binary frame as already discussed it is called as blob. However, for tracking direct use of these blobs is sensitive to noise and these leads to determine object identities as human at different times in a sequence.

Kalman filtering technique [8] is used for tracking object. Generally blob properties are the centroid, centroid velocity between consecutive frames, or size into a de-noised state trajectory. For tracking relevant blob these properties are used. The filter operates as a two-stage process switching between absorbing indication from the most recent foreground blob and creating a predictive distribution for the next. It is very essential to use correct Kalman filter in the context of tracking multiple objects simultaneously. For incorrect filter algorithm ends with wrong information.

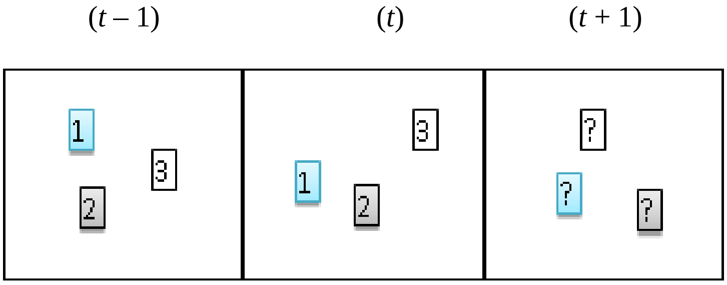

Figure 5: The Data Association Problem. The Object Identities at Time T + 1Must be Extrapolated while Being Robust against Missing Data and Noise [7]

The second part of the system is mainly concerned with developing appropriate framework. Tracking algorithm of this system is represented in Figure 5. Depending on previous frames *(t-1)* Observed statistics, the approximate identities of objects in next coming frame *(t+1)* must be predicted.

In tracking process when concern object is to be searched in consecutive frames data association problem arises between blob and Kalman filter while being robust against temporary occlusions and noise. It means additional complications arise if objects can temporarily disappear behind any wall or if non-existent foreground objects are observed. Significantly linear assignment problem is used, which optimizes an objective function that is the summation of all matching costs.

In mathematical equation set of pairwise costs was represented by square cost matrix *E*for which a permutation matrix U, u_{ij} Є {0,1} Was desired so that the summed cost is minimal. [7]

$$c(E, U) = \sum_{ij} e_{ij} \, u_{ij}$$

While finding various matching objects the cost function becomes helpful to make ensure that an optimal assignment maximizes the combined probability of all pairings. On the chosen object state, the chosen cost function is depends. And experiments were carried out to demonstrate encodings that can inform various association tasks.

Already developed algorithm compute the Linear Assignment Problem it as matching between set of same size to solve, for example Jonker and Volgenant's method [8]. When the number of available observations does not equal the number of Kalman filters, for the reference noisy observations or temporary object occlusions, it became problematic. For the instance when actual object measurements are lacking and only detection noise is available the tracking algorithm would sometimes be forced to take doubtful matching’s.

Two transformations of a cost matrix *E* were developed, to solve absent measurement and unlikely matching problems [7].

- To handle bogus or absent measurements problem, a rectangular cost matrix *R* can be augmented to a square matrix *E* by extending with a constant, but otherwise arbitrary dummy value ε_d. It was shown that the permutation matrix *U* that solves the Linear Assignment Problem on *E* directly induces a minimal-cost matching for the rectangular problem *R* for any constant ε_d.
- After that to ease the one-to-one matching constraints a $(n \times n)$ Cost matrix can be further extended to a $(2n \times 2n)$ Matrix F by adding three $(n \times n)$ Blocks of dummy values ε_d . It was shown that the permutation matrix V that solves the Linear Assignment Problem on F induces pairwise matching costs not exceeding ε_d. The parameter ε_d is thus the maximum acceptable matching cost, thereby letting us to reduce the matching constraints on the original problem R .

While properly handling the momentary object occlusions and extra noise problems need arise to combines these two methods. Then the data association problem was phrased in terms of readily available algorithms.

HUMAN POSE ESTIMATION

Final stage of paper describes a technique for human pose estimation in static images sequences based on a novel representation of part models. Pose estimation algorithm does not use articulated limb parts, but rather capture orientation with a mixture of templates for each part. It means flexible mixture model for capturing appropriate co-occurrence associations between parts, augmenting standard spring models that encode three-dimensional relations. The last matter must concerns initialization and the recovery from tracking failures. The filter can effectively search the entire statistics without a good algorithm for pose of human, because of the large number of unknown state parameter variables.

Discriminative Method for Pose Estimation

Finally a system uses Discriminative methods intention to recover pose directly from a set of measurements of blob from recent frame. From a single frame to a set of measurements of blob through some form of regression applied.

A set of training exemplars are

$$\mathcal{D} = \left\{ (S^{(i)}, Z^{(i)}) \sim p(S, Z) \middle| i = 1 \cdots N \right\}$$

And which are the joint distribution over states and measurements. The aim is for inputs, $Z \in \mathbb{R}^M$, are generic blob measurements from video sequences, and predict the outputs $S \in \mathbb{R}^N$ Represent the 3D poses of the human body. The simplest Discriminative method is Nearest-Neighbor (NN) [9], [10] where, specified a set of features detected in an image, the exemplar from the training database with the closest features is found, i.e $k^* = Arg\min_k D(\check{z}, z(k))$. The pose S^{k^*} for that exemplar is returned. One such approach was proposed by Shakhnarovich, Viola, and Darrell [11].

CONCLUSIONS

The algorithm is developed and evaluated on a number of different video sequences to revile both its strengths and weaknesses. Under varying conditions foreground object detection can produce useful information, especially when data is recoded in a different color space, as previous experiments directed. Manual assessment of various tracking results mentioned that multiple object tracking using the devised association technique performs adequately in many situations. Example tracking snapshots for one sequence were previously presented in Figure 1. In spite of the overall success, due to the limited representational power of foreground objects algorithm frequently fails to track several objects when one is briery occluded by another. The fundamentals of modern methodologies to pose estimation discussed here. Using the probabilistic formulation easily any one can built framework for tracking relatively simple motions of single isolated subjects in a compliant environment.

FURTHER WORK

The failure to track multiple objects that interact by occlusion highlights the blob tracking frame-work as fundamentally unsuitable for rational about occlusions. It is more useful to think about these interactions in terms of their causes namely an object occluding another rather than unpicking a complex mixture of their effects i.e. several objects merging into one. About this constraint, a reproductive model for image formation was adapted from Reid and Connor [2] and added to the object tracking algorithm.

The model preserves a sprite for each object that describes its status and can explain the materialization of an analyzed image by properly occluding a background reference image with different objects sprites. It is expected that by building a more complete model of how objects act together and occlude in the real world, future tracking improvements will be facilitated.

The more general problem of tracking arbitrary motion in monocular image sequences of unconstrained environments remains a puzzling and active area of research. While many improvements have been made, and the advancement is promising, no system to date can robustly deal with all the complications of improving the human pose and motion in a totally general situation.

REFERENCES

1. Ross Cutler. Larry S. Davis. 2000. Robust Real-Time Periodic Motion Detection, Analysis, and Applications. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. (August 2000) Volume 22. No. 8. 781-796.

2. Ismail Haritaoglu. David Harwood and Larry S. Davis. 2000. W^4: Real-Time Surveillance of people and their Activities. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence. (August 2000)Volume 22. No. 8. 809-830.

3. N. Dalal and B. Triggs. 2005. Histograms of oriented gradients for human Detection. Proceedings of the Conference on Computer Vision and Pattern Recognition. San Diego. California. USA. Pp. 886–893.

4. Q. Zhu. S. Avidan. M-C Yeh. K-W Cheng. 2006. Fast Human Detection Using a Cascade of Histograms of Oriented Gradients. Proceedings of the IEEE Computer Society Conference on Computer vision and Pattern Recognition. ISSN: 1063-6919. (June 2006).Volume 2. Pp. 1491-1498.

5. Chris Stauffer and W. Eric L. Grimson. 2000 Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(8):747-757.

6. I.D. Reid and K.R. Connor. 2005. Multiview segmentation and tracking of dynamic occluding layers. In BMVC 2005.

7. Fabian Wauthier. Motion Tracking. University of California. Berkeley.

8. R. Jonker and A. Volgenant. 1987. A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing. 384:325.

9. Howe N. 2007. Silhouette lookup for monocular 3d pose tracking. Image and Vision Computing 25:331–34.

10. Mori G. Malik J. 2002. Estimating human body configurations using shape context matching. In: IEEE European Conference on Computer Vision. 666–680.

11. Shakhnarovich G. Viola. Darrell TJ. 2003. Fast pose estimation with parameter sensitive hashing. In: IEEE International Conference on Computer Vision.750–757.

12. Kanaujia A. Sminchisescu C. Metaxas D. 2007a. Semi-supervised hierarchical models for 3D human pose reconstruction. In: IEEE Conference on Computer Vision and Pattern Recognition.

13. Marcus A. Brubaker. Leonid Sigal and David J. Fleet. Video-Based People Tracking.