Analysis of Human Categorization in Video Sequences

[1]Vaka Yamini ,[2]Bandla Srinivasa Rao
[1]PG Scholar, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP.
[2]Associate Professor & HOD, Dept. of CSE, VRS & YRN College of Engg. & Technology, Chirala, AP

Abstract: Visualization of moving objects is the most effective process in present days. Various techniques were introduced for tracking those type objects efficiently. This research has certain practical value and can be widely applied in some places like bank, prison, courtroom, large public facility large store house and military base. In video streaming process applications human detection is the main controversial issue in present days. The aim on moving human detection is to segment the regions corresponding to people from the rest of an image sequence. Considering these study analysis of the human detection we develop efficient techniques of human detection. This paper presents a set of techniques integrated into a low-cost PC based real time visual surveillance system for simultaneously human motion detection, tracking people, and analysis their activities in monochromatic video. The experimental result has shown the method present in this paper having the superiority of simple algorithm, fast recognition speed and high accuracy rate. Meanwhile, this method has certain robustness.

Index Terms: Human motion detection, object detection and classification, Subtraction method, Tread tracking system.

I. INTRODUCTION

Monitor zing human activity is one of the most important visual tasks to be carried out in visual surveillance scenarios. This task includes processes like target tracking, human activity characterization and recognition, etc. Human activity characterization and recognition is a special topic that has been addressed in the literature from different points of views and for different purposes. the objective was to characterize, aimed at building a feature representation for further recognition, the human activity of people in typical visual surveillance scenarios, like airport lounges, public building halls, commercial centers, etc., with a great variety of human action types and ordinary, rather poor, imaging conditions.

The recognition of texture and object categories is one of the most challenging problems in computer vision. Representation, detection and learning are the main issues that need to be tackled in designing a visual system for recognizing object categories. Interest point detection is an important research area in the field of image processing and computer vision. Image retrieval and object categorization heavily rely on interest point detection from which local image descriptors are computed for image and object matching. Color plays an important role in the pre-attentive stage in which features are detected as it is one of the elementary stimulus features. It is customary to define texture as a visual pattern characterized by the repetition of a few basic primitives.

ISSN-2320-7884 (Online)
ISSN-2321-0257 (Print)

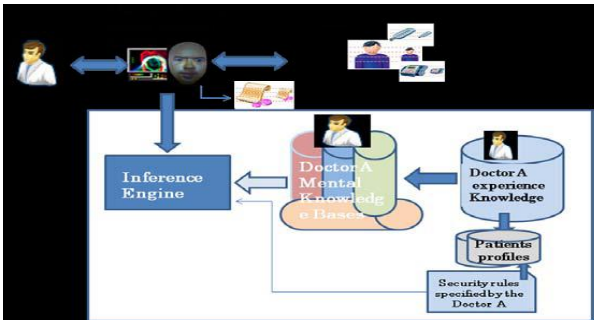

Figure 1: Human Motion Detection process for distributed video sequencing.

There is broad agreement on the issue of representation: object categories are represented as collection of features, each part has a distinctive appearance and spatial position. The current trend in object recognition is toward increasing the number of points applying several detectors or combining them or making the interest point distribution as dense as possible. With the explosive growth of image and video data sets, clustering and offline training of features become less feasible. By reducing the number of features and working with a predictable number of sparse features, larger image data sets can be processed in less time.

A stable number of features lead to a more predictable workload for such tasks. Recent work has aimed to find distinctive features by performing an evaluation of all features within the data set or per image class and choosing the most frequent ones. This approach requires an additional calculation step with an inherent demand on memory and processing time dependent on the number of features. This alternative may therefore provide selective search for robust features reducing the total number of interest points used for image retrieval. We propose color interest points to obtain a sparse image representation. Hence, we reduce the sensitivity to imaging conditions, light-invariant interest points are proposed. For color boosted points, the aim is to exploit color statistics derived from the occurrence probability of colors. Color boosted points are obtained through saliency-based feature selection. The use of color information allows extracting repeatable and scale-invariant interest points.

Color derivatives were taken to form the basis of a color saliency boosting function to equal the information content and saliency of a given color occurrence. Our aim is to select interest points based on color discriminative and invariant properties derived from local neighborhoods. Our focus is on color models that have useful perceptual and invariant properties to achieve a reduction in the number of interest points.

II. BACK GROUND WORK

The background subtraction method [6] is currently a common method for motion object detection; it is a technique which computes difference between the current image and the background model image and carries through threshold value processing to detect motion object.

1.1. Detection Motion Object

The motion object detection is the first step of all the object detection and tracking system meanwhile is the foundation of motion tracking and recognition. The quality of processed outcome in this phase directly influences the following effect, because in the course of the following processing we only consider the pixels corresponding to motion areas in images, so the function of motion detection in human body motion analysis is very important.

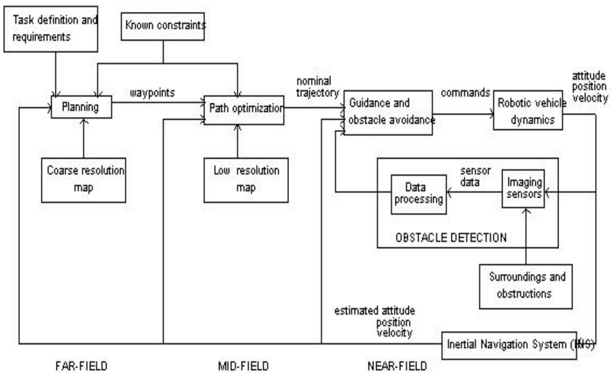

Figure 2: Motion detection of each credential formation.

The detection of motion areas is often influenced by light variation background variation and other background motion objects, thus selecting a steady and reliable detection method is very significant. In this paper we adopt the background subtraction method to detect motion object. So the system (VDS) and the MD are working together in comprehensive coherency; the former is complementary to the latter but not vice versa. The analysis is projected using his/her observed styles as a person and also as expert in medical diagnosis related practices. So there are different style of categorized knowledge reflecting such representation and related reasoning.

III. COMPONENTS OF REPRESENTATION

We first discuss scale- and affine-invariant local regions and the descriptors of their appearance, and then describe different image signatures and similarity measures suitable for comparing them. We use two complementary local region detector types to extract salient image structures:

- The *Harris-Laplace* detector It responds to corner-like regions
- The *Laplacian* detector It extracts blob-like regions

These two detectors are invariant to scale transformations alone as shown in the fig.2. We can either use rotationally invariant descriptors to achieve rotation invariance. The dominant gradient orientation is computed as the average of all gradient orientations in the region. We obtain affine-invariant versions of the Harris-Laplace and Laplacian detectors through the use of an *affine adaptation* procedure. Normalization leaves a rotational ambiguity that can be eliminated either by using rotation-invariant descriptors or by finding the dominant gradient orientation.

IV. APPROACH

An object model consists of a number of parts. Each part has an appearance, relative scale and can be occluded or not. The shape is represented by the mutual position of the parts. Entire model is generative and probabilistic shape and occlusions are all modeled by probability density functions. The process of learning an object category is one of first detecting regions and their scales and then estimating the parameters of the above densities from these regions.

This method finds regions that are salient over both location and scale. Each point on the image a histogram is made of the intensities in a circular region of radius (scale). The entropy of this histogram is then calculated and the local maxima are candidate scales for the region. The N regions with highest saliency over the image provide the features for learning and recognition.

Learning is carried out using the expectation maximization (EM) algorithm which iteratively converges from some random initial value of θ to a maximum. The scale information from each feature allows us to learn the model shape in a scale-invariant space. Learning complex models such as these has certain difficulties. Surprisingly, we assume given the complexity of the search space, the algorithm is remarkable consistent in its convergence. Initial conditions were chosen randomly within a sensible range and convergence usually occurred within 50-100 EM iterations.

V. HUMAN MOTION TRACKING

It is assumed in this study that the regions can enter and exit the scene and they can also get occluded by other regions. Regons carry informations like shape and size of the silhouette, and colors data on a bounding box location estimated for each person. Each region is defined by the 2D coordinates of the centroid, P, a ratio between the total number of foreground pixels (T) and the size of the bounding box (B), R = T/B, and the color/gray level characteristic, D. The regions, for which correspondence has been established, have also an associated velocity, V.

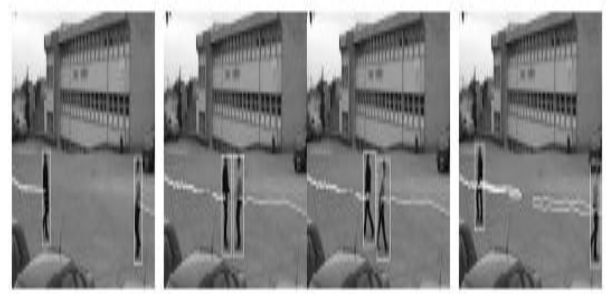

Figure 3: Tracking results on our data set (frames 2555-2606). Two person moving toward to each other, turn around each other, then go back away.

As shown in figure 3, two people moves toward each other. When they are occluded, they turn around each other, then go back away. The thick and thin white lines separately represent the trajectories of each object. The stored data in the trajectory map is used to implement each foreground region motions in the scene. When any foreground object in tracking is hidden to any nonmotion area (like passing at the behind of parked car), or temporarily occluded by other foreground object as they pass, the detected data of that object may not be obtained at the low level processing. At that or similar situations, high level implementation procedure is activated to estimate the possible position of that object using the previous tracked data obtained from trajectory. At the following frames, if the low level data about it is not obtained, its confidence is reduced. If the confidence of that object drops below a given threshold, it is considered lost, and is dropped from the tracking list stored in the trajectory map.

VI. HUMAN MOTION ANALYSIS

For the human motion analysis, using the geometrical shape models has the advantage that they have much information than directly obtained features. But the difficulty and cost of calculation in extracting the models from the input frames are disadvantage of using shape models for real time video surveillance applications. Those difficulties prevent researches from concentrating on cognition part of motion analysis process. Consequently, an approach depends on the variations of the features produced from the silhouette motions in frames is presented in this study. The features implemented for human motion analysis are directly obtained from dynamic variations on the star skeleton structures of the silhouette shape.

The three local extremal points correspond to head, and two legs. A human is moving in an upright position, it can be assumed that the uppermost skeleton segment represents the torso, and the lower segments determined by two extremal points represent two legs. Then the angle θ measures between the upper-most extremal point and vertical, the angle α measures between the lower two extremal points, and the angle β also measures moving variations in time between end locations of two extremal points in 2D space corresponding to the ankles, as shown in figure 4.a. (xc,yc) is the centroid of the motion blob (silhouette of the object under tracking).

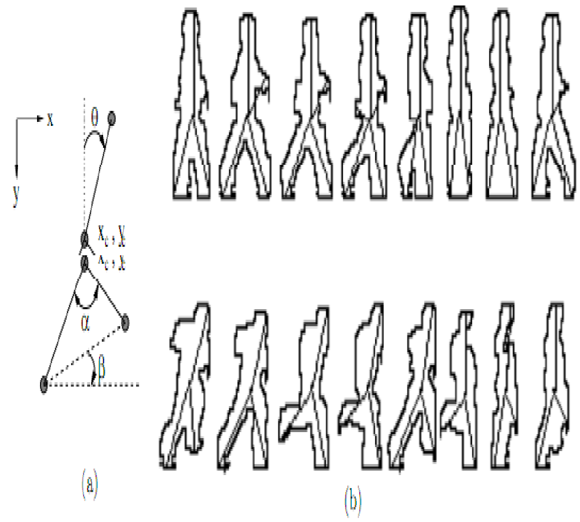

Figure 4: (a) Determining of posture features from the skeleton, (b) Silhouette and skeleton motion sequences of a walking and running person, respectively.

An approach to distinguish the walking and running actions was developed and tested on the different test sequences in our database. The most important features for distinguishing walking and running person can be produced by moving types of the foots in time, the characteristics on the foot cyclic and their speed variations. That features can be easily and simply obtained by manipulating the star skeleton properties.

VII. RESULT ANALYSIS

Experiments were carried out as follows: each dataset was split randomly into two separate sets of equal size. The spotted cat dataset was only 100 images originally and another 100 were added by reflecting the original images making 200 in total. There were two phases of experiments. Datasets with scale variability were normalized so that the objects were of uniform size. Algorithm was then evaluated on the datasets and compared to other approaches. The algorithm was run on the datasets containing scale variation and the performance compared to the scale-normalized case. The only parameter that was adjusted at all in all the following experiments was the scale over which features were found. The face and motorbike datasets have tight shape models but some of the parts have a highly variable appearance.

The majority of errors are a result of the object receiving insufficient coverage from the feature detector. One possibility is that the threshold is imposed on *N* many features on the object are removed. The feature detector seems to perform badly when the object is much darker than the background. The clustering of salient points into features within the feature detector.

The clustering of salient points into features within the feature detector. A recall- precision curve1 (RPC) and a table comparing the algorithm to previous approaches to object class recognition.

These mappings are expressed with various relationships between classes in the two different ontology fragments, for machine executable medical diagnosis purposes for interoperability based VDS.

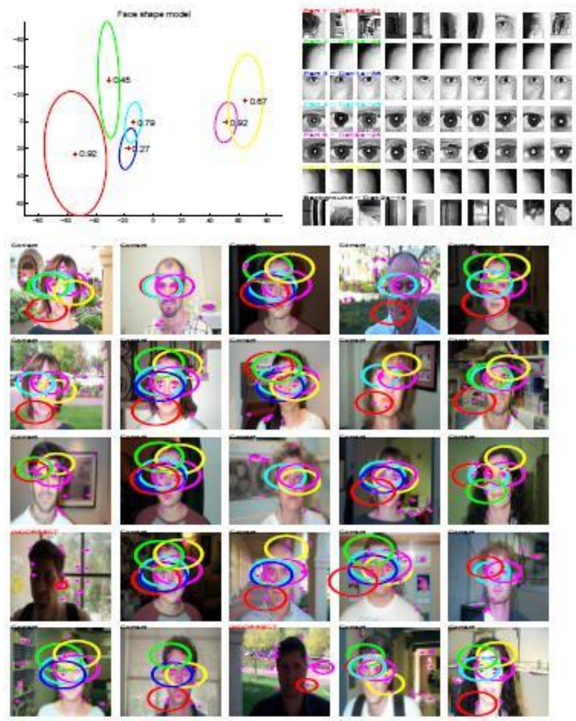

Figure 3: A typical face model with 6 parts

Size of the circles indicates the score of the hypothesis. Exactly the same algorithm settings are used for next consider example. As we consider the typical airplane with 6 parts.

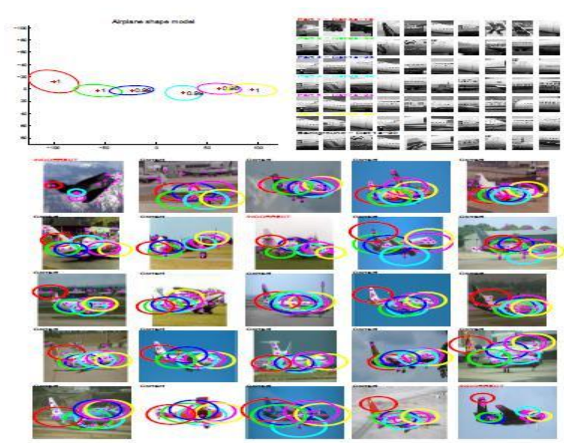

Figure 4: A typical airplane model in six parts.

The top left figure shows the shape model. Ellipses represent the variance of each part and the probability of each part being present is shown just to the right of the mean. Top right figure shows 10 patches closest to the mean of the appearance density for each part and the background density. Along with the determinant of the variance matrix, so as to give an idea as to the relative tightness of each distribution. The pink dots are features found on each image and the colored circles indicate the features of the best hypothesis in the image.

A video surveillance database is established for our experimental results. The database mainly contains video sequences on different days in outdoor and indoor environments. A digital camera (Sony DCR-TRV355E) fixed on a tripod and a CCD camera fixed on a pan-tilt motor platform are used to capture the video sequences. The algorithm for human motion detection, tracking and analysis presented in this paper has been implemented in C++ and runs under Windows 2K operating system at 96/133 MByte/MHz RAM, 850 MHz Celeron PC without using any special hardware. Currently, for 240 x 180 resolution gray-scale image sequences, the algorithm code without optimizing runs at 13-22 fps depending on the number of people in its field of view. Tests were performed on several sequences (each at least 30 minutes or more) representative of situations which might be commonly encountered in surveillance video.

VIII. CONCLUSION

In order to detect foreground objects, first, background scene model is statistically learned even the background is not completely stationary. A background maintenance model is also proposed for preventing some kind of falsies. Then, the candidate foreground regions are detected using thresholding, noise cleaning and their boundaries extracted using morphological filters. For human motion detection, object detection and classification approach for distinguishing a person, a group of person from detected foreground objects (e.g., cars) using silhouette shape and periodic motion cues is performed. Finally, the trajectory of the people in motion and several motion parameters produced from the cyclic motion of silhouette of the object under tracking are implemented for analyzing people activities such as walking and running, in the video sequences. Experimental results on the different test image sequences demonstrate that the proposed algorithm has an encouraging real-time background modeling based human motion detection and analysis performance with relatively robust and low computational cost.

IX. REFERENCES

[1] Eyup Gedikli, Murat Ekinci, “Human Motion Detection, Tracking and Analysis For Automated Surveillance”, Proc. of IEEE Vol. 89. No.10, (2001).

[2] R.T. Collins, A. J. Lipton, H. Fujiyoshi, T. Kanade.: Algorithms for Cooperative Multi sensor Surveillance. Proc. of IEEE Vol. 89. No.10, (2001).

[3] I. Haritaoglu, D. Harwood, L.S. Davis.: W4: Real-Time Surveillance of People and Their Activities. IEEE Trans. on PAMI, Vol. 22, No.8, (2000).

[4] H. Fujiyoshi, A. J. Lipton, T. Kanade.: Real-Time Human Motion Analysis by Image Skeletonization.

IEICE Trans. Inf.& SYST., Vol.E87-D, No.1, pp.113-120, January 2004.

[5] P. Perez, C. Hue, J. Vermaak, M. Gangnet, Color-Based Probabilistic Tracking, Proc. of European Conference on Computer Vision, Copenhagen, 27 May- 2 June 2002, Denmark.

[6] I. Haritaoglu, M. Flickner, Detection and Tracking of Shopping Groups in Stores, Proceeding of the 2001 IEEE Computer Vision and Pattern Recognition, Vol. 1, 8-14 December, 2001.

[7] Porikli F., Tuzel O., Human Body Tracking by Adaptive Background Models and Mean-Shift Analysis, IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, March, 2003.

[8] Mubarak Shah, Understanding human behavior from motion imagery, Machine Vision and Applications, Special Issue: Human modeling, analysis, and synthesis, Vol. 14, Issue 4, pp. 210-214, September 2003.

[9] A. K. Jain, A. Ross, S. Prabhakar, An Introduction to Biometric Recognition, IEEE Transactions on Circuit and Sytems for Video Technology, Special Issue on Imageand Video-Based Biometrics, Vol. 14, No.1, pp. 4-20, January 2004.

[10] M. Ekinci, F. W. Gibbs, B. T. Thomas.: Knowledge-Based Navigation for Autonomous Road Vehicles. Turkish Journal of Electrical Engineering and Computer, Vol. 8, No. 1, 2000.

[11] HAMIDO FUJITA, JUN HAKURA, MASAKI KUREMATSU, "**Human Interaction Based Reasoning Using Ontology Alignment**", Proceedings Of The 9th WSEAS International Conference On APPLICATIONS Of COMPUTER ENGINEERING.

[12] **Ye Zhang, Zhi-Jing Liu, " Irregular Behavior Recognition Based On Treading Track",** Proceedings Of The 2007 International Conference On Wavelet Analysis And Pattern Recognition, Beijing, China, 2-4 Nov. 2007.