HUMAN ACTION RECOGNITION USING BACKGROUND SUBTRACTION METHOD

P. Kalaivani[1*], Dr. S. Vimala[2#]

[1] *Ph.D Scholar, Department of Computer Science, Mother Teresa Women's University, TamilNadu, India*
[2]*Associate Professor, Department of Computer Science, Mother Teresa Women's University, TamilNadu, India*
** vanivijay2012@gmail.com*
 # vimalaharini@gmail.com

Abstract - *This paper addresses the view-independent action recognition from a different perspective using two web cameras. In the geometry-based methods we require identification of body parts and the estimation of corresponding points between video sequences. Differently to the previous view-based methods assume multi-view action samples for training and for testing. In the recent year, for image and video matching we explored local self-similarity descriptors. The main focus of our survey is on the human action recognition. In this survey we are discussed about the following action hierarchy Background subtraction, which has the Background representation, Classification, Background updating and Background initialization processes and finally, we presents the conclusion. This survey gives us a brief study about the action recognition of human under view changes using a static camera.*

Key Words: Background subtraction, self-similarities, classification, human action recognition.

1. INTRODUCTION

Visual recognition and understanding of human actions have attracted much attention over the past three decades and remain an active research area of computer vision. A good solution to the problem holds a yet unexplored potential for many applications, such as the search for and the structuring of large video archives, video surveillance, human-computer interaction, gesture recognition, and video editing.

Recent work has demonstrated the difficulty of the problem associated with the large variation of human action data due to the individual variations of people in expression, posture, motion, and clothing, perspective effects and camera motions, illumination variations, occlusions and disocclusions, and distracting effects of scenes surroundings. Also, actions frequently involve and depend on manipulated objects, which add another layer of variability. Most of the current methods for action recognition are designed for limited view variations. A reliable and a generic action recognition system, however, have to be robust to camera parameters and different viewpoints while observing an action sequence.

A database of poses seen from multiple view points has been created in Ahmad and Lee[1]. The multiview action recognition from a different perspective and avoids many assumptions of previous methods. Differently from the previous view-based methods, this does not assume multiview action samples either for training or for testing. This project approach builds upon self-similarities of action sequences over time.

In this paper we present a survey of Background subtraction, Pose estimation, Recognition, action primitives, actions, and activities and the conclusion of action recognition. This survey gives us a brief study about the action recognition of human under view changes. As the name suggests, background subtraction is the process of separating out foreground objects from the background in a sequence of video frames. Background subtraction is used in many emerging video applications, such as video surveillance, traffic monitoring, and gesture recognition for human-machine interfaces and so on. Many methods exist for background subtraction, each with different strengths and weaknesses in terms of performance and computational requirements. Pose estimation refers to the process of estimating the configuration of the underlying kinematic or skeletal articulation structure of a person.

The approach of recognition depends on the goal of the researcher and applications for activity recognition are interesting for surveillance, medical studies and rehabilitation, robotics, video indexing, and animation for film and games. For example, in scene interpretation the knowledge is often represented statistically and is meant to distinguish "regular" from "irregular" activities. The Action primitives will be used for atomic entities out of which actions are built. Actions are decomposed of several different activities. What do we mean by an action? Webster's dictionary defines action that doing of something, state of being in motion, the way of moving organs of the body, the moving of parts, guns, piano, military combat, appearance of animation in a painting,

sculpture, etc. More or less, hand gestures, sign language, facial expressions and lips movement during speech also the human activities like walking, running, jumping, jogging, etc, and aerobic exercises are all actions.

2. BACKGROUND SUBTRACTION METHOD

Background subtraction (BS) is a common and widely used technique for generating a foreground mask (namely, a binary image containing the pixels belonging to moving objects in the scene) by using static cameras. As the name suggests, BS calculates the foreground mask performing a subtraction between the current frame and a background model, containing the static part of the scene or, more in general, everything that can be considered as background given the characteristics of the observed scene.

Fig -1: Detection on multiple moving backgrounds (a) Original Image (b) MoG.

A. Background Representation

Background modeling consists of two main steps they are Background Initialization, and Background Update. In the first step, an initial model of the background is computed, while in the second step that model is updated in order to adapt to possible changes in the scene.

The MoG representation can be in RGB space, but also other color spaces can be applied, see [3] for an overview. Often a representation where the color and intensities are separated is applied, e.g., YUV, HSV, and normalized RGB, since this allows for detecting shadow-pixels wrongly classified as object-pixels [4]. Using a MoG in a 3D color space corresponds to ellipsoids or spheres (depending on the assumptions on the covariance matrix) of the Gaussian representations [2]. Other geometric representations are truncated cylinders [5] and truncated cones [6]. Conceptually different representations have also been developed. Elgammal et al. [7] use a kernel-based approach where they represent a background pixel by the individual pixels of the last N frames.

Haritaoglu et al. [8] represent the minimum and maximum value together with the maximum allowed change of the value in two consecutive frames. Heikkila and Pietikainen [9] represent each background pixel by a bit sequence, where each bit reflects whether the value of a neighboring pixel is above or below the pixel of interest, i.e., a texture operator. This makes the background model invariant to monotonic illumination changes. Oliver et al. [10] also use a pixel's neighbors to represent it.

They apply an Eigen space representation of the background and detect new objects by comparing the input image with an image reconstructed via the Eigen space. Eng et al. [11,12] divide a background model learnt over time into a number of non-overlapping blocks. The pixels within each block are grouped into at most three classes according to homogeneity. The means of these classes are then the representation of the background for this block, i.e., a spatio-temporal representation.

Heikkila and Pietikainen [9] have also applied their texture operator for a spatio-temporal block-based (overlapping blocks) background segmentation. Other spatio-temporal approaches are [13] and [14] where the background is represented by a predicted region found by an autoregressive process.

The choice of representation is not only dependant on the accuracy but also on the speed of the implementation and the application. This makes sense since the overall accuracy of background subtraction is a combination of representation, classification, updating, and initialization. For example, Cucchiara et al. [15] use only one value to represent each background pixel, but still good results (and speed) can be obtained due to advanced classification and updating. It should however be noted that the MoG representation is by far the most widely used method. For scenes with dynamic background the MoG representation does not suffice and methods directly aimed at modeling dynamic background should be applied, see e.g., [13,14].

B. Classification

A number of false positives and negatives will often be present after a background subtraction, for example due to shadows [4]. Using standard filtering techniques based on connected component analysis, size, median filter, morphology, and proximity can improve the result [7,15,16]. Alternatively, the fact that neighboring pixels are likely to be both foreground and background can be used in classification. Markov Random fields have been applied to implement this idea.

Recent methods have tried to directly identify the incorrect pixels and use classifiers to separate the pixels into a number of sub-classes: unchanged background, changes due to auto iris, shadows, highlights, moving object, cast shadow from moving object, ghost object (false positive), ghost shadow, etc. [15]. Classifiers have been based on color, gradients, flow information [15], and hysteresis thresholding [11].

C. Background Updating

In outdoor scenes, in particular, the value of a background pixel will change over time and an update mechanism is therefore required. The slow changes in the scene can be updated recursively by including the current pixel value into the model as a weighted combination [2,7,15]. A different approach is to measure the overall average change in the scene compared to the expected background and use this to update the model [6]. If no real-time requirements are present, both past and future values can be used to update the background. In general, for a good model update only pixels classified as unchanged background should be updated.

Rapid changes in the scene are accommodated by adding a new mode to the model. For the MoG model is a new Gaussian distribution, which is initiated whenever a non-background pixel is detected. The more pixels (over time) that support this distribution the more weight it will have. A similar approach is seen in [5,6] where the background model, denoted a codebook, for each pixel is represented by a number of codewords (cylinders or cones [4,5] in RGB-space). During run-time each foreground pixel creates a new codeword. A codeword not having any pixels assigned to it for a certain number of frames is eliminated. A similar idea can be found in [9].

D. Background Initialization

Initializing a background model requires robust statistical methods as the task should be robust against random occurrences of foreground objects, as well as against general image noise. A background model needs to be learned during an initialization phase. Earlier approaches assumed that no moving objects are present in a number of consecutive frames and then learn the model parameters in this period. However, in real scenarios this assumption will be invalid and recent methods have therefore focused on initialization in the presence of moving objects. In the MoG representation moving objects can to some extend be accepted during initialization since each foreground object will be represented by its own distribution which is likely to have a low weight. However, this erroneous distribution is likely to produce false positives in the classification process.

A different approach is to find only pixels that are true background pixels and then only apply these for initialization. This can be done using a temporal median filter if less than 50% of the values belong to foreground objects [7, 11]. Eng et al. [11] combine this with a skin detector to find and remove humans from the training images. Recent alternatives first divide the pixels in the initialization phase into temporal subintervals with similar values. Second, the ”best” subinterval belonging to the background is found as the subinterval with the minimum average motion (measured by optical flow) or the subinterval with the maximum ratio between the number of samples in the subinterval and their variance. The codeword method mentioned above uses a temporal filter after the initialization phase to eliminate any codeword that has not recurred for a long period of time [5]. A similar approach has used in [9].

3. HUMAN ACTION RECOGNITION

The field of action and activity representation and recognition is relatively old, yet still immature. This area is presently subject to intense investigation which is also reflected by the large number of different ideas and approaches. In scene interpretation, the representations should be independent from the objects causing the activity and thus are usually not meant to distinguish explicitly, e.g., cars from humans.

On the other hand, some surveillance applications focus explicitly on human activities and the interactions between humans. Here, one finds both, holistic approaches, that take into account the entire human body without considering particular body parts, and local approaches. Most holistic approaches attempt to identify “holistic” information such as gender, identity, or simple actions like walking or running.

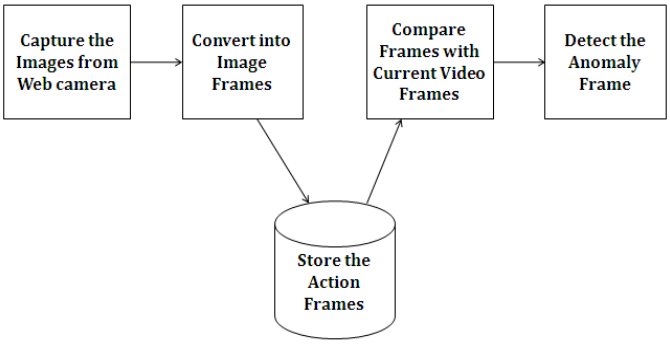

Fig -2: Anomaly Human Action Detection

This system focuses towards the security issues. Several researchers have proposed state of the art surveillance systems to help with some of the security issues with varying success. Recent studies have suggested that the ability of these surveillance systems to learn common environmental behavior patterns as wells as to detect and predict unusual, or anomalous, activities based on those learnt patterns are possible improvements to those systems.

4. ANOMALY DETECTION

Algorithm

1) The video to be detected is divided into segment. Here segment the continuous video sequence ‘V’

into 'N' segments

$V = \{V_1, V_2, V_3, \ldots\ldots Vn\}$,

2) Each segment is divided into frames. Event is detected for each frame.
A video segment Vn consists of Tn image frames.

$Vn = \{Im_1, Im_2, Im_3 \ldots . Im_{Tn}\}$.

3) The event is compared with the behavior model. Here Background model stores the values of a particular pixel which corresponds to the background colors. Pixel Change History (PCH) is represented for a pixel. Similar foreground pixels are grouped to form a blob. A behavior pattern is represented as a sequence of various events.

4) Build training data set group training behavior patterns upon which a model for normal behavior can be built.

5) If the event not exists already in training video, it is considered as anomaly.

4. CONCLUSIONS

In this paper we have analyzed the background subtraction approaches and human action recognition. The research in visual analysis of human movement must address a number of open problems to satisfy the common requirements of potential applications for reliable automatic tracking, reconstruction and recognition in future. Body part detectors which are invariant to viewpoint, body shape, and clothing are required to achieve reliable tracking and pose estimation in cluttered natural scenes. The use of learnt models of pose and motion are currently restricted to specific movements using the static camera.

ACKNOWLEDGEMENT

The Authors expresses their sincere thanks to Meenakshi College of Engineering, Chennai for providing necessary facilities to conduct the research work.

REFERENCES

[1] Ahmad.M and Lee.S, HMM-Based Human Action Recognition Using Multiview Image Sequences, vol. 1, pp. 263-266, 2006.

[2] C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, in: Computer Vision and Pattern Recogni-tion, Santa Barbara, CA, June 1998.

[3] F. Kristensen, P. Nilsson, V. Owall, Background segmentation beyond RGB, in: Asian Conference on Computer Vision, LNCS 3852, Hyderabad, India, Jan 13–16, 2006.

[4] A. Prati, I. Mikic´, M.M. Trivedi, R. Cucchiara, Detecting moving shadows: algorithms and evaluation, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (7) (2003) 918–923.

[5] K. Kim, T.H. Chalidabhongse, D. Harwood, L. Davis, Real-time foreground-background segmentation using codebook model, Real-Time Imaging 11 (3) (2005) 172–185.

[6] P. Fihl, R. Corlin, S. Park, T.B. Moeslund, M.M. Trivedi, Tracking of individuals in very long video sequences, in: International Symposium on Visual Computing (ISVC), Stateline, Lake Tahoe, Nevada, USA, Nov 6–8, 2006.

[7] A. Elgammal, D. Harwood, L. Davis, Non-parametric model for background subtraction, in: European Conference on Computer Vision, Dublin, Ireland, June 2000.

[8] I. Haritaoglu, D. Harwood, L.S. Davis, W4: real-time surveillance of people and their activities, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 809–830.

[9] M. Heikkila, M. Pietikainen, A texture-based method for modeling the background and detecting moving objects, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (4) (2006) 657–662.

[10] N. Oliver, B. Rosario, A. Pentland, A Bayesian computer vision system for modeling human interactions, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8) (2000) 831–843.

[11] H.L. Eng, K.A. Toh, A.H. Kam, J. Wang, W.Y. Yau, An automatic drowning detection surveillance system for challenging outdoor pool environments, in: International Conference on Computer Vision, Nice, France, Oct 13–16, 2003.

[12] H.L. Eng, J. Wang, A.H.K.S. Wah, W.Y. Yau, Robust human detection within a highly dynamic aquatic environment in real time, IEEE Transactions on Image Processing 15 (6) (2006) 1583–1600.

[13] A. Monnet, A. Mittal, N. Paragios, V. Ramesh, Background modeling and subtraction of dynamic scenes, in: International Conference on Computer Vision, Nice, France, Oct 13–16, 2003.

[14] J. Zhong, S. Sclaroff, Segmentating foreground objects from a dynamic textured background via robust kalman filter, in: Interna-tional Conference on Computer Vision, Nice, France, Oct 13–16, 2003.

[15] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, Detecting moving objects, ghosts, and shadows in video streams, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (10) (2003) 1337–1342.

[16] T. Zhao, R. Nevatia, Tracking multiple humans in complex situations, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (9) (2004) 1208–1221.

