ISSN (Online) : 2319 - 8753
ISSN (Print) : 2347 - 6710

International Journal of Innovative Research in Science, Engineering and Technology

Volume 3, Special Issue 3, March 2014

2014 International Conference on Innovations in Engineering and Technology (ICIET'14)

On 21st & 22nd March Organized by

K.L.N. College of Engineering, Madurai, Tamil Nadu, India

Discern Of Gestures and Tracking of Human Using Kalman Filter

S. Kanagamalliga[1], Dr. S. Vasuki[2], R.Sundaramoorthy[3], J.Allen Deva Priyam[4], M.Karthick[5]

[1]Assistant Professor, Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai ,India

[2]Professor Head , Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai, India

[3]U.G Student, Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai, India

[4]U.G Student, Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai, India

[5]U.G Student, Department of Electronics and Communication Engineering, Velammal College of Engineering and Technology, Madurai, India

ABSTRACT--This paper provides the interrelated topics of action recognition and detection of Humans. The foreground clutter is used to segment the human action from the video using the statistical method of Adaptive Background Mixture Model. The descriptor can be computed from drawing the boundary box for the humans in the video and the counting of actions is also displayed. The values for features are calculated from the descriptor. The descriptor allows for the comparison of the underlying dynamics of two space-time video segments irrespective of spatial appearance, such as differences induced by clothing, and with robustness to clutter. The calculated feature values are taken for extraction of human actions. An associated similarity measure is introduced that admits efficient exhaustive search for an action template, derived from a single exemplar video, across candidate video sequences also under occlusive conditions the human is detected using the adaptive filter.

KEYWORDS—Action detection, Action Representation, Foreground, Descriptor,Adaptive Background Mixture Model, Color Features, Kalman Filter

I. INTRODUCTION

A. Motivation:

This paper addresses the interrelated topics of detectingand localizing space-time patterns in a video. Specifically, patterns ofcurrent concern are those induced by human actions. Here, "action" refers to a simple dynamic pattern executed by an actor over a short duration of time (e.g., walking and hand waving). In contrast, activities can be considered as compositions of actions, sequentially, in parallel, or both. Potential applications of the presented research include video indexing and browsing, surveillance, visually guided interfaces, and tracking initialization.. Action detection seeks to detect and spatiotemporally localize an action, represented by a small video clip (i.e., thequery), within a larger video that may contain a large corpus of unknown actions. In the present work, action spotting is achieved by a single query video that defines the action template, rather than a training set of (positive and negative).

A key challenge in action detection arises from the fact that the same action-related pattern dynamics can yield very different image intensities due to spatial appearance differences, as with changes in clothing.Another challenge arises in natural imaging conditionswhere scene clutter requires the ability to distinguishrelevant pattern information from distractions.Clutter can be of two types:

1) Background clutter arises when actions are depicted in front of complicated, possibly dynamic, backdrops and 2) foreground clutter arises when actions are depicted with distractions superimposed, as with dynamic lighting, pseudo transparency (e.g., walking behind a chain-link fence), temporal aliasing, and weather effects (e.g., rain and snow). It is proposed that the choice of representation is key to meeting these challenges: Arepresentation that is invariant to purely spatial patternallows actions to be recognized independent of actorappearance; a representation that supports fine delineations of space-time structure makes it possible to tease action information from clutter. Also, for real-world applications such as video retrieval from the web, computational efficiency is a further requirement.

B. Related work:

A wealth of work has considered the analysis of human actions from visual data, e.g., [1],[2].One manner of organizing this literature is interms of the underlying representation of actions. A brief corresponding survey of representative approaches follows: Tracking-based methods begin by tracking body parts,joints, or both and classify actions based on features extracted from the motion trajectories, e.g.,[3],[4],[5], [6]. General impedimentsto fully automated operation include tracker initialization and robustness .Consequently, much of this work has been realized with some degree of human intervention.

Other method shave classified actions detection and tracking based on features extracted from color histograms body shape sasrepresented by contours or silhouettes, with the motivation thatsuch representation sarero bust to spatial appearance details This class of approach relies on figure-ground segmentation across space-time,with the drawback tha trobust segmentation remain selusive in uncontrolled settings .Further,silhouette do not provide information on the human body limbs when they are infront of the body(i.e.,inside thesilhouette) and thus yield ambiguous information.

II. ACTION DETECTION DATASETS

The Weizmann dataset consists of 10 action categories with 9 people, resulting in 90 videos. In the Weizmann dataset, a static and simple background is used throughout the videos. Simple human actions of `running', `walking', `jumping-jack', `jumping forward on two legs', `jumping in place on two legs', `galloping sideways', `waving one hand', `waving two hand', and bending' are performed by the actors. The resolutions of the videos are 180*144, 25fps. Both dataset are composed of relatively simple action-level activities, and only one participant appears in the scene. What we must note is that these datasets are designed to verify the `classification' ability of the systems on simple actions. Each video of the datasets contains executions of only one simple action, performed by a single actor. That is, an entire motion-related feature extracted from each video corresponds to a single action, and the goal is to identify the label of the video while knowing that the video belongs to one of a limited number of known action classes. Further, all actions in both datasets except for the `bend' action of the Weizmann dataset are periodic actions (e.g. walking), makingthe videos suitable for action-level classification systems. The Action detection datasets are used in various research and scholarly articles for video processing the below dataset is a widely used dataset in the video processing research since it is widely accepted by all the research areas.

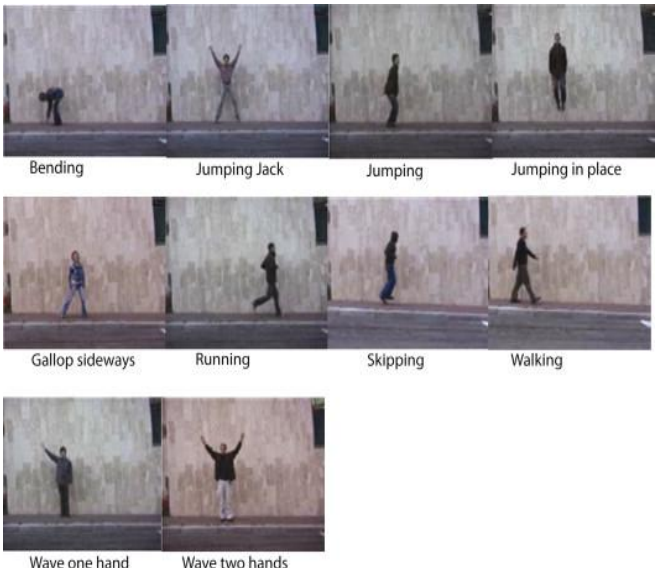


Fig 1.Weizmann datasets

Further, they are particularly suitable for recognition of periodic actions;spatio-temporal features can be extracted repeatedly from the periodic actions. Figure 1 compares the classification accuracies of the systems. The X axis corresponds to the time of the publication, while the Y axis shows the classification performance of the systems. Most of the systems tested on the Weizmann dataset have obtained successful results, mainly because of the simplicity of the dataset. Particularly, [Blank et al. 2005; Niebles et al. 2006; Rodriguez et al. 2008; Bregonzioet al. 2009] have obtained more than 0.95 classification accuracy.

III. STATISTICAL METHOD FOR FOREGROUND

Adaptive background mixture model is used for extraction, namely, segmenting a human body or objects from a background, is usually the first and enabling step for many high-level vision analysis tasks, such as video surveillance, people tracking and activity recognition [5].Once a background model is established humans in the video frames can be detected as the difference between the current video frame and the background model. In this work, as a part of our on-going research effort to develop automated video-based activity monitoring analysis for eldercare, we propose an accurate and robust background extraction and human tracking algorithm which is capable of operating in real-world, unconstrained environments with complex and dynamic backgrounds. In our eldercare research project, we use video cameras to collect information about elderly residents' daily extract important activity information to perform automated functional assessment and detect abnormal events, such as a person falling onto the floor.

Fig 2 .Adaptive mixture model segmentation for human actions

IV. BOUNDINGBOX REPRESENTATIONS

Incorporating the person bounding box into the classifier: Previous work on object classification [7] demonstrated that background is often correlated with objects in the image (e.g. cars often appear on streets) and can provide useful signal for the classifier. The goal here is to investigate different ways of incorporating the background information into the classifier for actions in still images. We consider the following three approaches:

A. "Person": Images are centered on the person performing the action, cropped to contain1:5_ the size of the bounding box and re-sized such that the larger dimension is 300pixels. This setup is similar to that of Gupta et al. , i.e. the person occupies the majority of the image and the background is largely suppressed.

B1. "Person-Background': The original images are resized so that the maximum dimension of the 1:5_ rescaled person bounding box is 300 pixels, but no cropping is performed. The 1:5_ rescaled person bounding box is then used in both training and test to localize the person in the image and provides a coarse segmentation of the image into foreground (inside the rescaled person bounding box) and background (the rest of the image). The foreground and background regions are treated separately. The final kernel value between two images X and Y represented using foreground histograms xf and yf , and background histograms xb and yb, respectively, is given as the sum of the two kernels, K(x;y) = Kf (xf ;yf)+Kb(xb;yb).

B2. "Person-Image": This setup is similar to B1, however, instead of the background region, 2-level spatial pyramid representation of the entire image is used. Note that approaches A, B1 and B2 use the manually provided person bounding boxes at both the training and test time to localize the person performing the action. This simulates the case of a perfectly working person detector.

V. COLOR FEATURES

Human vision system is more sensitive to color information than grey levels so color is the first candidate used for feature extraction. Color histogram is one common method used to represent color contents. The algorithm follows a similar procession: selection of a color space, representing of color features, and matching algorithms. It is the most common one used for images on computer because the computer display is using the combination of the primary colors (R,G,B) to display any perceived color. Each pixel in the screen is composed of 3 points which is stimulated by R,G,B electron gun separately. However RGB space is not perceptually uniform so the color distance in RGB color space does not correspond to color dissimilarity in perception. Therefore we prefer to transform image data in RGB color space to other perceptual uniform space before feature extraction

VI. TRACKING USING KALMAN FILTER

A Kalman filter is an optimal recursive data processing algorithm The Kalman filter incorporates all information that can be provided to it. It processes all available measurements, regardless of their precision, to estimate the current value of the variables of interest computationally efficient due to its recursive structure Assumes that variables being estimated are time dependent

VII. Implementation And Simulation Results

Let us consider the input as video sequence from Weizmann datasets which is related to single human activity (like avi or mpeg format).From that, we evaluate the performance for silhouette extraction, bounding box representations by simulations conducted in MATLAB

A. Input video

Let us consider the input as Weizmann datasets which it perform human activity like bend, jump, walk, run etc.....where, we had taken two activities as bend and jump. Bend activity consist of 84 frames. Similarly walk activity consists of 67 frames.

Fig 3. Consider a single human activity for single video. (a) Human activity like walk.

B. Background segmentation

For single human activity recognition silhouette segmentation is accurate and robust. It is one of the background modelling. It can be used under Gaussian mixture model. Let $I(x,y)$be a new image entered into a system, a reconstructed image.$I^{'}(x,y)t$is obtainedusing M eigenvectors information. A moving object,$D_i(x,y)$is a difference between$I(x,y)$ and $I^{'}(x,y)t$as the following equation:

$$D_i(x,y) = |I(x,y) - I^{'}(x,y)t|$$

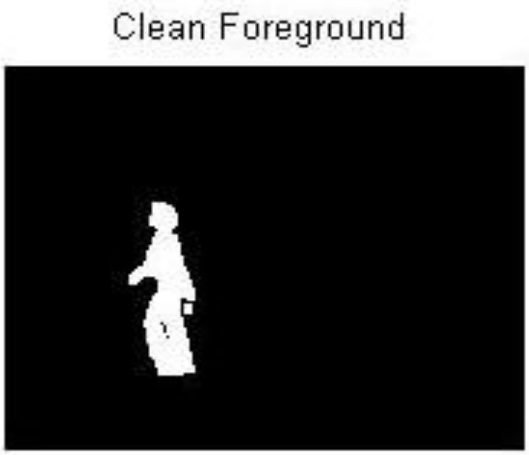


Fig. 4Background segmentation for walk activity

C. Bounding box representation

We represent an activity video as a composition of its scene background and the foreground videos of the sub events composing the activity. Sub-events are atomic-level actions (e.g. stretching an arm, withdrawing an arm, and moving forward), for example, pushing activity is composed of multiple sub events which it includes one person stretching an arm, other person pushed away.

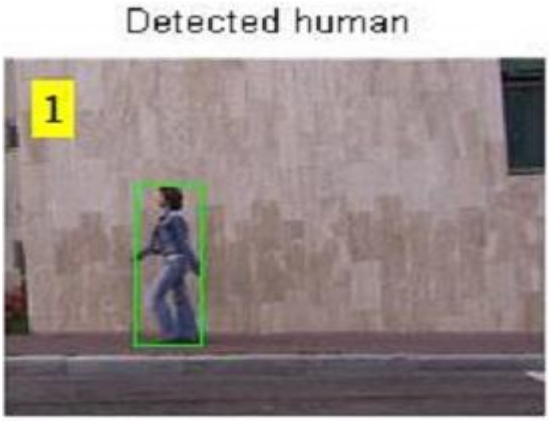


Fig 5 Bounding Box representation

Formally, we represent an activity video in V by its three components V = (b,G,S), Where b is the background image, G describes the spatial location of the activity's centre, where G=(c ,d ,o).S is the sub set of events,$s = (s_1, s_2 \ldots . s_i)$where$s_i$is the sub event. Each $s_i$contains four types of information.$s_i = (e_i, a_i, r_i, t_i)$. e_iIs the sequence of foregroundimages obtained during the sub-event, where is the length of the foreground sequences, $e_i = \left(e_i^0, e_i^1 \ldots . . e_i^{ni}\right)$.

Then,a_i indicates the actor id performing the sub-event, r_i is the normalized bounding box specifying the sub-event's spatial location, which is described relatively with respect to

$r_i = (r_i^{left}, r_i^{right}, r_i^{height}, r_i^{width})$Where$r_i$

Multiplied by d specifies the relative occurring region of thesub-event $s_i$in the video. Where, is a normalized time interval specifying the centre of the interval and its duration.Let the starting time of the sub event be $start_i$ and the ending time of it be end_i.Then

$$t_i^{loc} = \frac{start_i + end_i}{2.o} \quad t_i^{dur} = \frac{end_i - start_i}{o}$$

Furthermore, foreground videos of the sub-events, are maintained independently

D. Color Features

The idea of an intensity histogram can be generalized to continuous data, say the signals represented by real functions or images represented by functions with two-dimensional domain.

Let$f \in L^1(R^n)$(see Lebesgue space)

then the cumulative histogram operator H can be defined by:

$$H(f)(y) = \mu\{x: f(x) \leq y\}$$

μ is the Lebesgue measure of sets. H(f)In turn is a real function. The (non-cumulative) histogram is defined as its derivative.

$$h(f) = H(f')$$

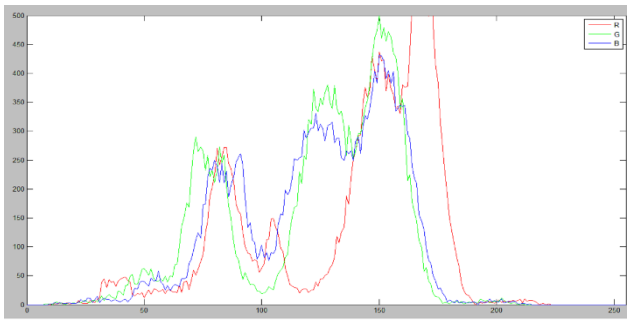
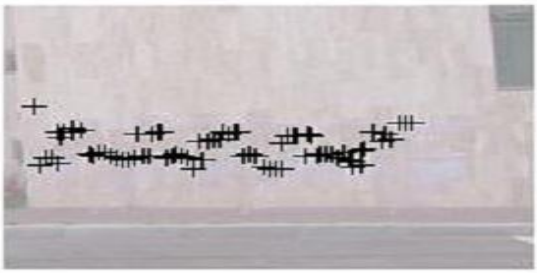


Fig. 6 Color Features in RGB Histogram

By comparing histograms signatures of two images and matching the color content of one image with the other, the color histogram is particularly well suited for the problem of recognizing an object of unknown position and rotation within a scene

E. Action Detection and Tracking of Human using Kalman Filter

In order to use the Kalman filter to estimate the internal state of a process given only a sequence of noisyobservations, one must model the process in accordance with the framework of the Kalman filter



Fig, 7 Detection of Human Action

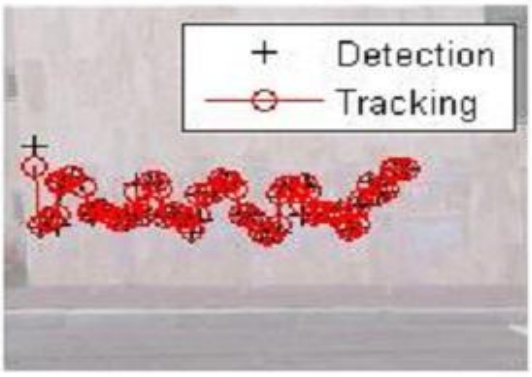


Fig. 8 Detection and Tracking of Human

This means specifying the following matrices: F_k, the state-transition model; H_k, the observation model; Q_k, the covarianceof the process noise; R_k, the covariance of the observation noise; and sometimes B_k, the control-input model, for each time-step, *k*, as described below.

The Kalman filter model assumes the true state at time *k* is evolved from the state at (*k*−1) according to

$$X_k = F_k X_{K-1} + B_K U_k + W_K$$

Where

- F_K is the state transition model which is applied to the previous state X_{K-1}
- B_K is the control-input model which is applied to the control vector U_K
- W_K is the process noise which is assumed to be drawn from a zero mean multivariate normal distribution with covariance Q_K

$$W_K \sim N(0, Q_K)$$

At time K an observation (or measurement) Z_K of the true state X_K is made according to

$$Z_K = H_K X_K + V_K$$

Where H_K is the observation model which maps the true state space into the observed space and Y_K is the observation noise which is assumed to be zero mean Gaussian white noise with covariance R_K

$$V_K \sim N(0, R_K)$$

The initial state, and the noise vectors at each step $\{X_0, W_1, \ldots . W_K V_1 \ldots \ldots V_k\}$ are all assumed to be mutually independent.

VIII. CONCLUSION

We have presented a new approach to detect the action and track the human using a single exemplar video. Adaptive background mixture model is used to segment the human from the background. Bounding box is drawn to calculate the values for feature extraction. RGB histogram values are taken to enhance the video and also to calculate the features. Finally, Kalman filter is used to detect the actions of human in the video and also to track the human.

REFERENCES

[1] P.Turaga, R.Chellappa, V. Subrahmanian, and O. Udrea, "Machine RecognitionofHumanActivities: A Survey," IEEE Trans. Circuits andSystemsforVideoTechnology, vol.18, no.11, pp.1473-1488,Nov.2008

[2] R.Poppe,"A Survey on Vision-BasedHuman Action Recognition,"ImageandVision Computing,vol.28,no.6,pp.976-990,2010.

[3] Y. Yacooband M.Black,"ParameterizedModelingand Recognition of Activities,"Computer Vision and Image Understanding, vol.73,no.2,pp. 232-247,1999.

[4] D.Ramananand D.Forsyth, "AutomaticAnnotationofEveryday Movements" proc.Neural Information Processing Systems 2003

[5] C.Fanti, L. ZelnikManor, and P. Perona,"HybridModels for HumanMotionRecognition,"Proc.IEEEConf.ComputerVisionand Pattern Recognition,pp. 1166-1173,2005.

[6]S.Ali,A.Basharat,and M.Shah,"ChaoticInvariantsforHuman Actions Recognition" Proc 11th IEEE Int'l Conf. computer vision 2007

[7] Y.Ke,R Sukthankarand M.Hebert, "Event Detectionin 2007.

[8] J. Yuan,Z. Liu,and Y.Wu,"DiscriminativeVideoPatternSearch for Efficient Action Detection,"IEEETrans.Pattern Analysisand MachineIntelligence,vol.33,no.9,pp.1728-1743,Sept.2011