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

A Robust Method to Detect Human Actions by Fusing Hierarchically Filtered Motion with Stip Features

P.Karpagavalli[#1], A.Alagusathya[*2], Dr.A.V.Ramprasad[#3]

Department of Electronics and communication Engineering , K.L.N College of Engineering, pottapalayam, Madurai ,Tamil Nadu, India.

Department of Electronics and communication Engineering , K.L.N College of Engineering, pottapalayam, Madurai ,Tamil Nadu, India

Department of Electronics and communication Engineering , K.L.N College of Engineering, pottapalayam, Madurai ,Tamil Nadu, India

ABSTRACT— In recent years the human actions recognition in crowd video for security purpose in airport, highways or roads, metro line station, street etc. In this paper we are describing the different human actions in crowd video. In existing system, filtering parameters are sensitive in complex scenes and detected interest points are heavily affected by the cluttered background. So in order to handle cluttered background we propose a hierarchical filtered motion (HFM) with spatiotemporal interest point features to recognize more type of actions in crowd. The interest points are detected by 2-D Harris corner points detector and MHI Hierarchical motion filter is used to reduce the distracting motion and also find the characterize points. Finally the different human actions are classified by the GMM. The proposed method can detect three different human actions like boxing, hand waving, clapping. The proposed method is analyzed and validated by using KTH dataset and MSR action dataset II.

KEYWORDS— crowd video, Hierarchical Filtered Motion (HFM), Motion History Image (MHI)

I. Introduction

Action recognition is difficult in cluttered background. Our aim of the work is to recognize human actions in crowd environments. By combining multiple features will rectify the action detection problems.

Recent work on action recognition that are based on Laptev et al.[3],[4]used local spatiotemporal invariant points(STIPs)[8] local spatio-temporal descriptors Histogram of Oriented Gradients(HOG)/ Histogram of Oriented Flows (HOF)[7],[9],and multichannel nonlinear SVMs for realistic actions in movies. Yuan et al. employed the same features (STIPs) and descriptors (HOG/HOF) and proposed a discriminative sub volume search for efficient action detection by the use of a nearest neighbor-based classifier.

II. System Overview

In this paper we propose a hierarchical filter motion is used to extract motion information and reduce distracting motions. The 2-D Harris corners and MHI for interest-point detection. To eliminate the distracting motions we are applied hierarchical filtered motion method in crowd. Finally it will classify by the GMM. The proposed approach achieves the results on the standard KTH dataset that has clean background. We perform action recognition on the MSR action dataset II that consists of three actions (hand clapping, hand waving, and boxing).

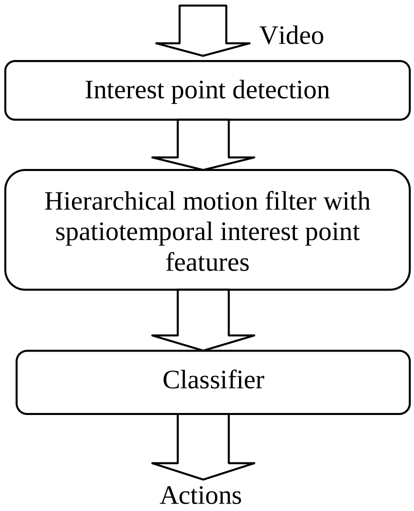


Fig. 1Framework of the proposed method

III. PROPOSED METHOD

INTEREST-POINT DETECTION

A. Motion History Image

The MHI is a static image template where pixel intensity is a function of the recency of motion in a sequence. By using the motion template technology to recognize human movements within interactive environments. Gradients in the MHI is better than optical flow are more efficient to compute.

B. 2-Dharris Corner Points Detection

It is a step of algorithms that rely on identifying characteristic points or interest pointsit The image intensity will change largely in multiple directions.This can alternatively be formulated by examining the changes of intensity due to shifts in a local window.From it will identify the feature points.

IV. HIERARCHICAL MOTION FILTER

A.Global Motion Filter

We are apply a motion smoothing step at the MGI to remove the isolated motion directions to get filtered motion field-smoothed gradients of the MHI. For local filtered motion field processing,we decompose the MHI as a number of layers with different motion directions.

B. Local Motion filter

we apply a local filtered motion field at each interest point between the pixels in the local region and the interest point.the local region is the window to calculate HOG-MHI.Let d(Po,B) denote the minimum distance between po and all the points in B.

SPATIOTEMPORAL INTEREST POINT FEATURE

A novel spatio-temporal (ST) feature which is based on the SURF (Speeded-Up Robust Feature) features For designing a new ST feature,we use only moving interest points where ST features are extracted and discard static interest points, because we expect that it is a local feature which represents how objects in a video are moving. We calculate the variation between the each frames.This values will be used as a feature values of video.

============= (1)

V. FEATURE DISCRIPTORS

The window is divided into an (nx,ny) grid of patches. We use Histograms of Oriented Gradient (HOG) without considering the directions and Histograms of Oriented Gradient in Motion History Image (HOG_MHI)the action recognition is decreases without considering directions The multiscale process will heavily increase the size of the feature vector for training and testing.The size of each window is calculated by $w_x=k_{nx}$ and$w_y=k_{ny}$.we use randomly selected window sizes between w_{min} and $w_{max.}$

VI. ACTION CLASSIFICATION

GAUSSIAN MIXTURE MODEL

We employ a GMM is to have the ability to model any given probability distribution function when the number of mixture component is large.Given a K-component GMM,the probability of a patch x is,

$$P_r(x|\Omega) = \sum_{k=1}^{K} N(x: \mu_k \Sigma_k) \quad (2)$$

N ($x;\mu_k, \Sigma_k$) denotes the normal distribution with mean μ_k and variance $\sum_k$. The set of all the parameters of the GMM is denoted as $\Omega = N(w_k, \mu_k, \Sigma_k)$, $1 \le k \le K$.

Suppose there are C category corresponds to a GMM with K components $\Omega c = \{wk, \mu_k, \Sigma_k\}$. This can be solved by the EM algorithm that is an iteractive method between performing an expectation step (E-step) and maximization step (M-step).

$$p_{ik}^c = \frac{w_{kN}(x_i;\mu_k^0 \Sigma_k^0)}{\sum_{k'=1}^{k} w_{k'N}(x_i;\mu_{k'}^0,\Sigma_{k'}^0)} \; for\; x_i \epsilon X^c \quad (3)$$

After obtaining the GMM parameters $\Omega^1, \Omega^2, \ldots, \Omega^c$, we can easily classify a new video clip according to the action category.

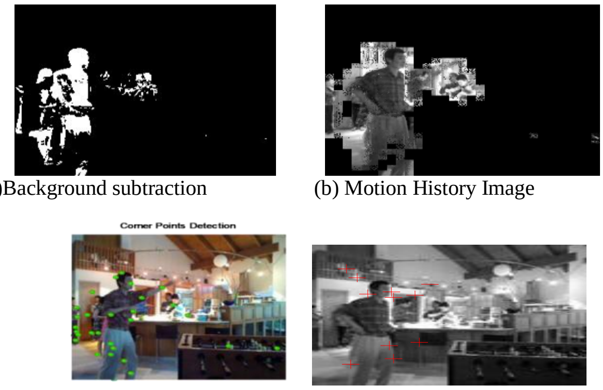
VII. EXPERIMENTS RESULTS AND DISCUSSION

The KTH dataset was used as standard benchmark for action recognition. It has four controlled environments with clean background (indoors, outdoors, outdoors with scale variation, and outdoors with different clothes.)

The current video database containing six types of human actions(walking, jogging, running, boxing, hand waving and hand clapping) performed several times by 25 subjects in four different scenarios: outdoors *s1*, outdoors with scale variation *s2*, outdoors with different clothes *s3* and indoors *s4* as illustrated below. All sequences were taken over homogeneous backgrounds with a static camera with *25*fps frame rate. The sequences were down sampled to the spatial resolution of *160x120* pixels and have a length of four seconds in average.

The MSR action dataset II consists of three type of action like, boxing, hand clapping, hand waving, GMM is trained from the KTH dataset for six actions to recognize the actions on the MSR action dataset II. Figure 2 Shows the Interest points are detected by 2-D Harris corner detection, MHI calculation. Processing of both global and local filters and calculation of HOG and HOG-MHI.

MSR ACTION DATASET II EFFICIENCY ANALYSIS

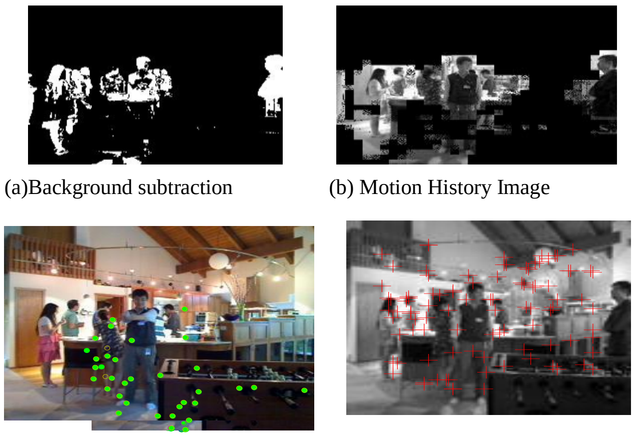


(a)Background subtraction (b) Motion History Image

(c) Corner points detection (d) Feature extraction

Fig. 2 the action recognition from Boxing

Figure 3 shows the background subtraction for separated foreground image and interest points are detected by 2-D Harris corner detection, MHI calculation and feature extraction



(a)Background subtraction (b) Motion History Image

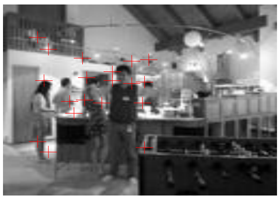
(c)Corner points detection (d) Feature extraction

Fig. 3The action recognition from Handclapping

Figure 4 shows the background subtraction for separated foreground image and interest points are detected by 2-D Harris corner detection, MHI calculation and feature extraction

(a)Background subtraction (b) Motion History image

(c)Corner points detection



(d) Feature extraction

Fig. 4 The action recognition from Hand waving

TABLE I

COMPARISON with RESULTS on the KTH ACTION DATASET

Method	Accuracy
Dollar et al[11]	80.7%
Yin et al.[12]	82%
Kaaniche et al.[13]	90.57%
2Dcorners+IpDetection+HOG/HOG- MHI	93.9%
2Dcorners+Ipdetection+Hierarchical motion filter +HOG/HOG-MHI	93.6%
STIP+HOG/HOF+GMM	94.5%

Precision and recall curve are defined as,

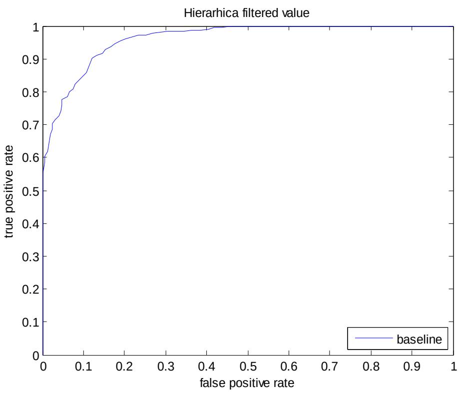
Recall: the ratio of true positives correctly classify such,

Precision: the ratio of true positives among the entire sample positive as negative.

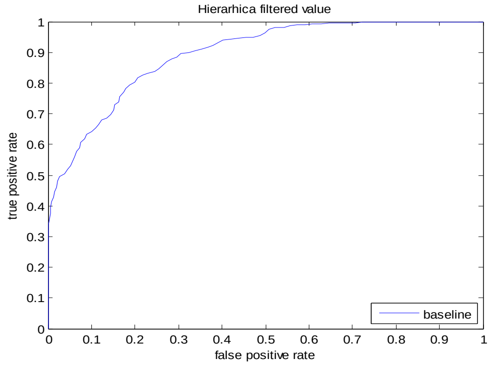
Where, true positive (TP): the sample is classified as being from the class and it classify is, according to the ground.

False negative (FN): the sample is actually positive, but is classified as negative.

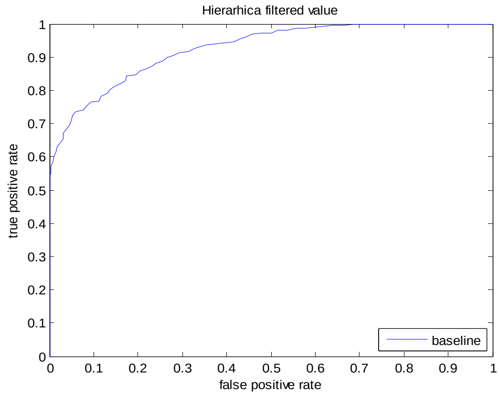
False positive (FP): the sample is wrongly classified as positive



(a) Boxing



(b) Handclapping



(c) Hand waving

Fig 5 shows the action detection by the use of HFM (a) Boxing, (b) Hand clapping, (c) Hand waving

VIII.CONCLUSION

We propose a novel method to detect the Action in the crowd video without tracking objects or key points. A Hierarchical Filtered Motion was proposed to reduce distracting motions caused by the background moving objects near an interest point. We have performed action classification and detection Experiments on videos with cluttered and moving background. We have performed HFM with spatiotemporal interest point features and branch –and-bound based detection.HFM feature based on the Motion history images (MHI), 2D harries corner. Finally the Action is classified by the GMM. Our proposed method shows the better performance than the existing system and also suitable for real time action recognition.

REFERENCES

[1] Y. Tian, L. Cao, Z. Liu, and Z. Zhang, Action Detection by Fusing Hierarchically Filtered Motion With Spatiotemporal Interest Point Features In book Human Behavior Recognition Technologies: Intelligent Applications for Monitoring and Security, by IGI Global, Hans Guesgen and Stephen Marsland (Eds.), ISBN 978-1-4666-3682-8 (hardcover) -- ISBN 978-1-4666-3683-5 (ebook), 2013.

[2] Yingli Tian, LiangLiang Cao, Zicheng Liu zhengyou zhangyou " Hierarchical filtered motion for Action recognition in Crowded videos"IEEE Trans on systems, vol, 42 n0.3 may 2012

[3] I. Laptev, M. Marszalek, C. Schmid, B. Rozenfeld, Learning Realistic Human Actions from Movies, CVPR, 2008.

[4] M. Marszałek, I. Laptev, and C. Schmid. Actions in context. In CVPR, 2009.

[5] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," in Proc. IEEE Conf. Comput. Vis., vol. 2, pp. 1395–1402, Oct. 2005.

[6] M. Blaschko and C. Lampert, "Learning to localize objects with structured output regression," in Proc. Eur. Conf. Comput. Vis., 2008, pp. 2–15

[7] L. Cao, Y. Tian, Z. Liu, B. Yao, Z. Zhang, and T. Huang, Action Detection using Multiple Spatial-Temporal Interest Point Features, IEEEInternational Conference on Multimedia & Expo **(**ICME), 2010

[8] I. Laptev and T. Lindeberg. Space-time interest points. In ICCV, 2003.

[9] Y. Pang, Y. Yuan, X. Li, and J. Pan, Efficient HOG human detection,Signal Processing, Vol. 91, Issue 4, Pages 773-781, 2011.

[10] A. F. Bobick and J.W. Davis, "The recognition of human movement using temporal templates," IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 3, Mar. 2001, pp. 257–267.

[11] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie,: Behavior recognition viaspatio-temporal features. In: Proc. of ICCV Int. work-shop onVisual Surveillance and Performance Evaluation of Tracking andSurveillance (VSPETS), pp. 65–72, 2005.

[12] J. Yin and Y. Meng, Human ActivityRecognition in Video using aHierarchical Probabilistic Latent Model, In CVPR, 2010.

[13] M. B. Kaaniche and F. Bremond, Gesture Recognition by LearningLocal Motion Signatures, In CVPR, 2010.

[14] H. Wang, M. Ullah, A. Klaser, I. Laptev, and C. Schmid, Evaluation oflocal spatio-temporal features for action recognition, BMVC, 2009.