Journal of Scientific & Industrial Research
Vol. 79, July 2020, pp. 626-630

A Feature Weighting Technique on SVM for Human Action Recognition

Soumya Ranjan Mishra[1]*, K Deepthi Krishna[2], Goutam Sanyal[1] and Anirban Sarkar[1]
[1]Department of Computer Science and Engineering, NIT Durgapur
[2]Department of Computer Science and Engineering, VMTW, JNTUH

Received 17August 2019; revised 22 March 2020; accepted 11 May 2020

Human action recognition is a challenging research topic and attracted very good attention in the last few years. This paper presents a features weighting framework for human action recognition based on the movement of different body-parts. Intuitively, Understanding the motion of a particular body-part having a major contribution to a specific action gives a better representation of that human activity. For example, action like walking, running and jogging, movement of the leg is more important and in boxing, waving and clapping, hand movement is more effective. This work presents a technique, utilizing the sub-region body-parts recognition rate to the weight kernel function. First, the complete human body is extracted from the background and HOG (histogram of gradient) based body-part detection is applied to generate three different sub-region (head, arm and body, foot and leg) of complete human body. Recognition rate and weight is calculated for all these sub-region (body-parts) for a particular action. Based on the weight (ω) of sub-region, a weighted feature Gaussian kernel function is obtained and weighted feature support vector machine (WF-SVM) classifier is constructed. The experimental results of the proposed framework have better performance on both KTH and UCF-ARG datasets compared against several state-of-the-art methods.

Keywords: Histogram of gradient, Interest point, Optical flow, Weighted feature SVM

Introduction

In the area of computer vision and machine learning, many researchers are working with the space-time pattern recognition where the object of interest is extracted and identified from the video sequence. We have proposed a new technique called weighted feature human action recognition. The proposed model is motivated by the fact that every human action can be described by the movement of body parts and different body parts having different impact on the action recognition result. Here in this work, we extracted features from different sub-region and calculate the impact and weight of each sub-region features based on the magnitude and recognition rate. In the recognition stage, the obtained weight result is used to calculate weight kernel function and we get a new human action recognition model based on SVM with weighted kernel function.

Proposed Model Description

The proposed model description is shown in Fig. 1. Moving human body detection and segmentation, Human body part detection with different sub-region, Motion features extraction (Optical flow), and action classification based on weighted feature SVM are basic objective of our system.

Human body Detection and Segmentation

Human action depends on body movement therefore we need to extract only the moving object (Human body) from the background which is significant part in action recognition. Visual attention technique can be used to segment moving object from background.[1] The moving objects detection follows the steps explained below.

Average and Gaussian filter is applied on video $I(x, y, t)$ of size $(N \times M)$ at time t, Then *saliency* value at each pixel position of image (x, y) is calculated as:

$$S(x,y) = d[I_{avg}\,(x,y), I_{gaussian}(x,y)]$$

Human Body part detection

Here we have used body-parts detectors to detect different parts of the human body (Head region, Arm and body region, Leg and foot region) based on sliding window technique.[2] Fixed size of rectangular bounding boxes is used for individual object region. These bounding boxes slide over the image and each sub window considers as an input to these body parts detectors. All these inputs are then independently classified as respective objects. All those detected

*Author for Correspondence
E-mail: soumyaranjanmishra.in@gmail.com

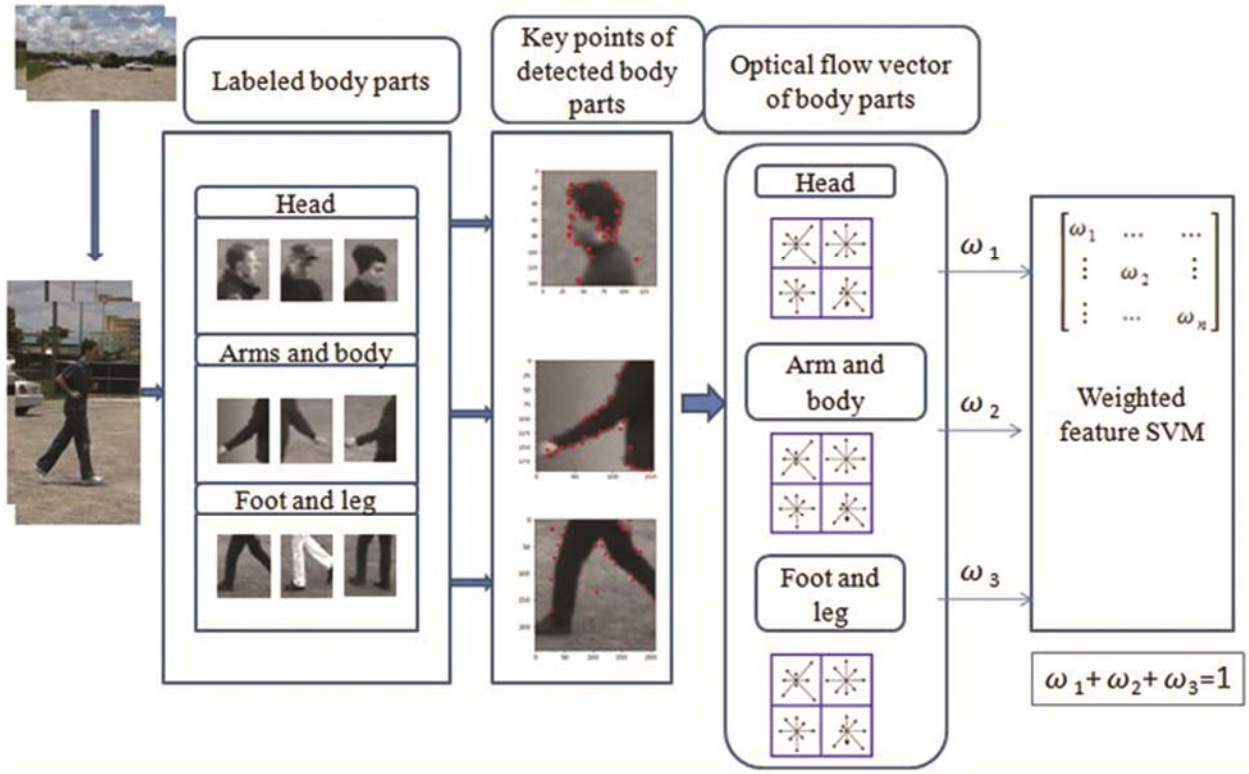


Fig.1 — Proposed weighted feature framework

body parts are combined into a proper geometrical shape into another fixed size bounding box as a complete human shape to prevent possible false positives.

In this approach, All training images of different body parts, divided into (8×8) cell and applied sobel operation and HOG features is extracted to form a normalized 6-bin histogram.[3] As a result, we got 6-dimension feature vector for each cells. We select the best feature group from all these feature vector based on minimization of the standard deviation (σ).

Algorithm to train three different body parts:

1. Apply Sobel operator for every training image (head, arms and legs).
2. Divide the Sobel image into 8 × 8 cells.
3. Calculate HOG features for each cell within the range $(-\pi/2, +\pi/2)$.
4. Store the values in a feature array.
5. Apply feature selection technique over the feature array.
6. Train the selected feature using SVM.

Interest point Detection from individual body parts

The above approach gives individual body parts as output from full human body. Now to model the movement of individual body parts, first we have to track some interesting key-points from all detected body parts and then apply optical flow (OF) to compute the movements of those points. We have used an Improved Corner Detection Algorithm[4] to find interest point and optical flow feature to model motion pattern of interest point in human action video.[5,6] Small image patches called as window is moved around the target images to get the variation in intensity in both the direction x and y. After this with each window, a score R is to be computed and with a threshold to this score, corners are selected.

Weighted Feature SVM

The optimal hyperplane is obtained by solving constraint optimization problem and get the classification decision function by introducing a kernel function K in a new high dimension space H as follows:

$$f(x) = sign\left[\sum_1^l a_i y_i k\left(\vec{x_i}\,\vec{x} + b\right)\right] \quad \ldots (1)$$

Weighted features kernel function k_p based on SVM is defined as:

$$k_p\left(\vec{x}_i, \vec{x}_j\right) = k\left(\vec{x}_i{}^T P, \vec{x}_j{}^T P\right) \quad \ldots (2)$$

Here P is referred to as a weighted feature matrix. The different value for P lead to different weight situation of different types of matrix below:

I: When P is an identity matrix of order n, this is no weight situation.

II: When P is an diagonal matrix of order n, Where $(P)_{ii} = w_i\ (1 \le i \le n)$ is the weight of i^{th} feature and not all weight w_i are equal.

III: When P is an arbitrary matrix of order n. This is the full weight situation.

For this work we have taken P as a diagonal matrix of order n, and the weighted features kernel function can be process from eq (1) for Gaussian kernel function[7,8] as follows:

$$k_p\left(\vec{x}_i, \vec{x}_j\right) = \exp\left(-\gamma \left\| \vec{x}_i{}^T P - \vec{x}_j{}^T \right\|^2\right) =$$

$$\exp\left(-\gamma\left(\left(\vec{x}_i - \vec{x}_j\right)^T PP^T\left(\vec{x}_i - \vec{x}_j\right)\right)\right) \quad \ldots (3)$$

Experiment

In this section, we evaluate the performance of the proposed WF-SVM using two benchmark dataset: UCF-ARG dataset, and KTH human action dataset. UCF Human action dataset consists of 10 action performed by 12 actors recorded from 3 different angles. We have taken the ground view recoding for our experiment. The 10 actions are Boxing, Clapping, Digging, Jogging, Carrying, Open-Close Trunk, Running, Throwing, Walking and Waving. KTH dataset consist of 6 actions (boxing, hand-clapping, hand-waving, jogging, running and walking) by 25 different people in different environments. The frame size is 960×540 and after background subtraction (moving object detection and segmentation), it is reduced to 64×128. In this paper we have presented a novel and efficient framework by modeling the motion of human body part. We have divided the 64×128 cropped image into three different region (Head, body and arms, and leg and foot) and extracted feature points of all three different region separately by using shi-tomasi interest point detection and then we iteratively track those points using lucas-kanade optical flow(OF) to obtain motion features of three different region.

1-Head region: From the head region key points are detected and computed the optical flow (OF) to obtain a motion feature vector which stores the displacement of these key points from first frame to second and so on.

2-Body and arm region: Key points are selected and optical flow (OF) vector is obtained.

3-Leg and foot: Similarly key points are selected from foot and leg region to generate optical flow (OF) vector.

We have extracted optical flow features (OF) from all three different region of human body from its current and previous frame for all different category action videos having N number of frame per video. We can extract a set of $N\text{-}1$ optical flow features descriptors per video. The optical flow (OF) descriptor size depends on the number of interest point detected per region multiply by 2. For example, if 40 key points are detected from head region then the optical flow descriptor size will be $40 \times 2 = 80$ (2 comes from horizontal and vertical direction) .We have extracted optical flow from 3 different region of human body for all videos in UCF dataset (Total video per action 48, so for 10 action we have used 480 video files). Then we split the extracted optical flow into training and testing set and stored separately (for training we have used 28 video per sample and 20 for testing. K-means clustering with a fixed number of clusters is used for vector quantization to assign optical flow descriptors to its nearest code-word. Finally we construct Bag-of-word (BOW) vector for each region of human body, BOW vector is like a histogram which counts the frequency of optical flow descriptors in video. We then train a weighted feature kernel function $K_p(f_i, f_j)$ SVM in our training set.

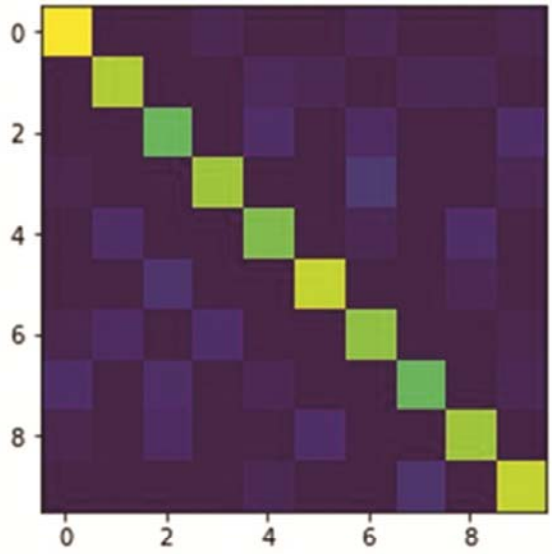
We experiment five times under different body region features for all 10 different actions and observed that, the influence of features of three different region of body are discrete in different action class. For example the legs and foot body parts are having major contribution for action like walking, jogging and running and arms for boxing, waving and clapping. Based on the analysis of our experiments in three features area we obtained the weight w_i of each body region for different action and corresponding linear transformation square matrix P below:

$$w = w_1 + w_2 + w_3$$

We can get the values of w_1, w_2 and w_3 from the recognition rate of different region of body for all 10 different actions.

ACTION	TEST SET	SVM	WF-SVM
Boxing	40	32	35
Carrying	40	28	31
Clapping	40	28	32
Digging	40	30	34
Jogging	40	30	34
Open Trunk	40	29	32
Running	40	34	36
Throwing	40	29	32
Walking	40	27	30
Waving	40	29	32

(a)



ACTION	TEST SET	SVM	WF-SVM
Boxing	32	25	28
Clapping	32	23	27
Waving	32	21	25
Jogging	32	25	29
Running	32	25	30
Walking	32	24	28

(b)

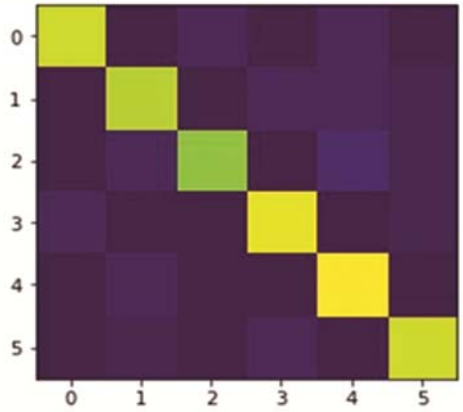


Fig. 2 — (a) Confusion matrix of UCF dataset by 40 random splits and the corresponding values under two kernel function; (b) Confusion matrix of KTH dataset by 32 random splits and the corresponding values under two kernel function

$$P = \begin{bmatrix} w_1 & . & . \\ . & w_2 & . \\ . & . & w_3 \end{bmatrix} = \begin{bmatrix} w_1 P_1 & 0 & 0 \\ 0 & w_2 P_2 & 0 \\ 0 & 0 & w_3 P_3 \end{bmatrix}$$

Weighted feature kernel function $K_p\left(\vec{f}_i, \vec{f}_j\right)$ obtained from eq(3) for different body parts feature vector $\vec{f}$

$$k_p\left(\vec{f}_i, \vec{f}_j\right) = \exp\left(-\gamma \left\| \vec{f}_i^{\,T} P - \vec{f}_j^{\,T} \right\|^2\right) =$$

$$\exp\left(-\gamma\left(\left(\vec{f}_i - \vec{f}_j\right)^T PP^T\left(\vec{f}_i - \vec{f}_j\right)\right)\right) \quad \ldots (4)$$

Action recognition performance

Finally, we report the performance of our proposed WF-SVM approach on UCF and KTH dataset. The average precision of WF-SVM which uses weighted feature Gaussian kernel on UCF dataset is 82%, which is better than standard SVM with linear kernel whose average precision is 74%. Similarly on KTH dataset, the average precisions are 86% by using WF-SVM and 78% by using SVM with linear kernel. The confusion matrix for UCF and KTH dataset of proposed approach is shown in Fig 2.

From these confusion matrixes it is evident that most of the actions are correctly classified by our proposed WF-SVM than standard SVM shown in tabular format in Fig. 2. The reason behind this is the reduction of the influence of weak correlation features by weighted feature. We have used python and open cv for feature extraction, K-means clustering is used with cluster size 200 for vector quantization to assign optical flow feature to its nearest codeword.TF-IDF weighting scheme is used for BOW (Bag of word).

Conclusions and Future Work

In this paper, a feature weighting technique is proposed for human action recognition, since the effect of each human body region on action recognition is different. The experimental results suggest the weighted features Gaussian kernel function has better performance in human action recognition on both KTH and UCF action dataset. This body part-based approach does not beat few best state-of-the-art methods on KTH and UCF dataset because of resolution and scaling of the video. Cropping the moving object and detecting individual body-parts perform better in video having high resolution and good indoor controlled environment. However the proposed approaches have shown impressive performance on controlled dataset in laboratory setting. To distinguish very similar action, more number of body parts and joints can be taken as sub-region and sub-region features ranking can be done for more accurate inter-class human action classification.

References

1 Patel C I, Garg S, Zaveri T, Banerjee A & Patel R, Human action recognition using fusion of features for unconstrained video sequences, *Comput Electr Eng*, **70** (2016) 284–301.

2 Jiang X, Pang Y, Pan J & Li X, Flexible sliding windows with adaptive pixel strides, *Signal Process*, **110** (2015) 37–45.

3 Chakraborty B, Bagdanov A D, Gonzalez J & Roca X, Human action recognition using an ensemble of body-part detectors, *Expert syst*, **30** (2013) 101–114.

4 Han S, Yu W, Yang H & Wan S, An Improved Corner Detection Algorithm Based on Harris, *CAC, Xian, China*, 2018, 1575–1580.

5 SevillaLara L, Liao Y, Guney F, Jampani V, Geiger A & Black M J, On the Integration of Optical Flow and Action Recognition in *Pattern Recognition*, by T Brox, A Bruhn, M Fritz (GCPR 2018, lecture notes in computer science, Springer, Cham), Vol. 11269, 2019.

6 Mishra S R, Mishra T K, Sarkar A & Sanyal G, PSO based combined kernel learning framework for recognition of first-person activity in a video, *EvoI Intell*, **12** (2018) 1–7.

7 Wei W & Jia Q, Weighted Feature Gaussian Kernel SVM for Emotion Recognition, *Comput Intell Neurosc* (2016) Volume 2016 |Article ID 7696035 | 7 pages | https://doi.org/10.1155/2016/7696035

8 Ding M & Fan G, Articulated and Generalized Gaussian Kernel Correlation for Human Pose Estimation, *IEEE Trans on Image Procrs* **25** (2016) 776–789.