Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

Action Change Detection in Video Based on HOG

M. Fakhredanesh**, *S. Roostaie

Faculty of Electrical and Computer, Malek Ashtar University of Technology, Tehran, Iran.

Article Info

Article History:
Received 07 March 2019
Reviewed 12 May 2019
Revised 17 July 2019
Accepted 10 December 2019

Keywords:
Artificial Intelligence
Computer vision
Machine learning
Video surveillance
Motion analysis

*Corresponding Author's Email Address:
m-fakhredanesh@aut.ac.ir

Abstract

Background and Objectives: Action recognition, as the processes of labeling an unknown action of a query video, is a challenging problem, due to the event complexity, variations in imaging conditions, and intra- and inter-individual action-variability. A number of solutions proposed to solve action recognition problem. Many of these frameworks suppose that each video sequence includes only one action class. Therefore, we need to break down a video sequence into sub-sequences, each containing only a single action class.

Methods: In this paper, we develop an unsupervised action change detection method to detect the time of actions change, without classifying the actions. In this method, a silhouette-based framework will be used for action representation. This representation uses xt patterns. The xt pattern is a selected frame of xty volume. This volume is achieved by rotating the traditional space-time volume and displacing its axes. In xty volume, each frame consists of two axes (x) and time (t), and y value specifies the frame number.

Results: To test the performance of the proposed method, we created 105 artificial videos using the Weizmann dataset, as well as time-continuous camera-captured video. The experiments have been conducted on this dataset. The precision of the proposed method was 98.13% and the recall was 100%.

Conclusion: The proposed unsupervised approach can detect action changes with a high precision. Therefore, it can be useful in combination with an action recognition method for designing an integrated action recognition system.

Introduction

Human operators cannot be able to manage the enormous volumes of videos that are generated by network cameras. Therefore, it is important to develop efficient and effectual automatic methods to analyze video data [1].

Action recognition problem is defined as follows: "given a dictionary of annotated training action videos, recognize the unknown action of a query video" [2]. In the other words, first, a dictionary of labeled training data, such as walking, running, and jumping videos are given. After an unknown action video is fed to the system, the action recognition part analyzes the unknown action based on the given videos and declares the action label.

Action recognition has various interesting applications in many areas. Some of these applications are indoor and outdoor video surveillance, wildlife monitoring, human-computer interaction [1], the human health system, sign language, virtual reality, and humanoid robot [3]. Despite a significant effort, the action

recognition is still a challenging problem. The main challenges in this area are variety and complexity of the events that may occur in a video (e.g., clutter, and occlusions), variations in the imaging conditions (e.g., illumination, viewpoint, and resolution) and different appearance from the same action that did by different people [1].

There are two basic components in action recognition: action representation and action classification.

Some successful action representations are based on motion models, shape models, interest point models, dynamic models and Geometric human body models.

In Motion models to distinguish between actions, dynamic characteristics of actions are important and motion features should be extracted. Motion models may be affected by background motion. Weng and Guan [4] proposed stacked trajectory energy image (STEI) method. In this method, trajectories are extracted from motion saliency regions. In their research, a three-stream CNN framework is proposed to simultaneously capture spatial, temporal and global motion information of the action from RGB frames, optical flow, and STEI. Wang et al. [5] proposed a method that utilized dense trajectories to describe actions. The main idea of their research is to densely sample feature points in each frame, and track them in the video based on optical flow.

In shape models, features are extracted from a shape, which is silhouette of the tracked object. A sequence of these silhouettes forms a silhouette tunnel, i.e., a spatiotemporal binary mask that it shows the deformation of the moving object over time. A silhouette tunnel is desirable for action recognition because of being insensitive to color, texture, and contrast changes. Silhouette tunnels are also known as object tunnels [6], [7], activity tubes [8] and space-time volumes (STVs)[9]. Vishwakarma and Kapoor [10] have used space-time volume to action recognition. Amraji et al. [11] presented a method in which the shape represented using Fourier Descriptors (FDs) as features. They used Principal Component Analysis (PCA) to project these features into Eigen-space. Then, the KNN classifier is used. In Bobick and Davis [12] research, the basis of the action representation is a temporal template. They proposed a motion energy image (MEI), that indicating the presence of motion, and motion history image (MHI), that is a scalar field depicting the recency of motion in a sequence. Sharif et al. [13] proposed a human action recognition method based on statistical weighted segmentation (SWS) and feature selection approach. Their proposed technique is comprised of two primary steps: a) Efficient human segmentation from video sequences, and b) Features extraction, fusion and finally selection of most robust features. The authors of [14] presented a method to recognize human actions. They used Histograms of Oriented Gradients (HOG) for human pose representations. In [15], human activities are recognized using background subtraction, HOG features and Back-Propagation Neural Network (BPNN) classifier. In this approach, background estimation is performed at first, using mean filter to obtain the background and areas of the image containing important information. Afterwards, in order to extract features to describe human motion, a histogram of oriented gradients (HOG) descriptor is used, with the idea that local shape information can be completely described by intensity gradients or edge directions. Finally, a BPNN is used to perform the final classification. Khan et al. [16] proposed a framework that primarily consolidated four phases: a) acquisition and preprocessing, b) frame segmentation which incorporates top-hat and bottom-hat filters along with the proposed RGB* color space enhancement, c) features extraction and dimensionality reduction, and d) classification using SVM classifier. Dalal and Triggs [17] used HOG descriptors for extracting feature sets to human detection and linear SVM for classification. The accuracy of shape models is highly related to the quality of silhouettes. Interest points, e.g., corners, etc., are sufficiently discriminative and are usually far fewer than the number of pixels in a video sequence. Zhu et al. [18], evaluated the spatio-temporal interest point (STIP) based features for depth-based action recognition. Niebles et al. [19] used 2D Gaussian and 1D Gabor filters to find local region of interest in the cuboids of space and time. Zhang et al.[20] , proposed manifold regularized Sparse Representation (MRSR). In this approach, each interest point is represented by its local closest words. MRSR has an analytical solution and is easy to calculate. Khan et al. [21], proposed a new method for action recognition by fusion of deep neural network (DNN) and multi view features. In this method, both types of features are fused in serial approach, and selection of best among them is done through three parameters i.e. relative entropy, mutual information, and strong correlation. Later, all these parameters are combined by employing a mean parallel fusion approach, and help in designing a high probability based threshold function to select the best features. These features are finally provided into Naïve Bayes classifiers for final recognition. Nazar et al. [22] have used CNN based features and classical features in a parallel processing. The best of them were selected before fusion stage. At the end, the labeled data is returned as an output by using classifier. Arshad et al. [23] proposed a technique that works in two phases: a) two pre-trained CNN models are applied and their information is fused via a parallel approach. In the proposed parallel fusion

approach, both feature vectors are compared with each other and a strongly correlated feature is added into a fused matrix. b) Entropy and skewness vectors are calculated from the fused matrix. The best subsets of picked features are finally fed to multiple classifiers and finest one is chosen based on accuracy value.

Dynamic model's general idea is to define each static posture of an action as a state, and describe the dynamics (temporal variations) of the action by using a state-space transition model. An action is modeled as a set of states and connections in the state space using a Dynamic Bayesian Network (DBN). Hidden Markov Model (HMM) is proven to be a special type of DBN with a fixed structure of inference graph to model time variations of data features [2] directly. The basic human interactions are modeled by coupled hidden Markov models (CHMMs) [24]. In this model, multiple state variables of CHMMs are temporally corresponded to the conditional probabilities of one chain given the other chain. The interval temporal Bayesian network (ITBN), a graphical model has been proposed by Zhang et al. [25] to model the temporal dependencies over time intervals. This model has been developed the Bayesian network by the interval algebra. In geometric human body model, the pose of human body must first be estimated to recognize human action using this approach. Although all these techniques have promising results, but they have a huge limitation, and it is the extraction of human body model and body joint points. The recent advanced cost-effective depth cameras help in the extraction of human body joint points, but depth cameras also have some limitations. First, the range of the depth sensor is limited; Second, skeleton tracking and the estimated 3D joint positions are noisy and can produce inaccurate results or even fails when serious occlusion occurs [26]. The human body modeling is done in 3D and 2D methods. The 3D methods [27, 28] generally perform better than 2D methods, because the 3D methods exploit all the views to evaluate a query action, unlike the reported 2D methods which are limited to a single-view testing. 3D methods usually have higher computational complexities than 2D models and they are not suitable for real-time applications [29]. Das et al. [30] have achieved good results by combining skeleton information and apparent features. Liu and Wang [31] proposed an action representation method named Part Movement Model (PMM), which captures the spatial-temporal structure of human actions and divides the actions into discriminative part movements. The algorithms used to classify human action so far, are generally be categorized as dynamic time warping (DTW), generative models, discriminative models and others such as Kalman filter, binary tree, Kernel-based and k-nearest neighbors (KNN). The dynamic time warping (DTW) [32] is a method for similarity measure in order to compare two temporal sequences. DTW is simple, but it is not appropriate for a large number of classes with many variations. Some generative models (dynamic classifiers) are proposed such as Hidden Markov Models (HMM) [3]. Shian-Ru Ke [33] uses three-dimensional modeling of the body and HMM classification algorithm. The discriminative models (static classifications) such as SVMs and artificial neural networks (ANNs) [34] can also be used at the action classification. Hoai et al. [35] proposed a method that simultaneously performs video segmentation and action recognition using a multi-class SVM framework and the dynamic programming. Vishwakarma and Rajiv [10] recognized actions using space-time and SVM-NN classification algorithms.

The performance of both generative and discriminative models relies on extensive training dataset. Therefore, other methods such as Kalman filter, binary tree, multidimensional indexing, and k-nearest neighbors (KNN) were proposed to comprise this problem [3]. As the result, different classification algorithms usually require different sets of suitable feature representations [2].

Many of frameworks that proposed to solve action recognition problem, suppose that each video sequence includes only one action class. Therefore, we need to break down a video sequence into sub-sequences, each containing only a single action class. In this paper, we focus on the specific problem of action change detection, i.e., identifying when one action stops and another action begin. Even if action representation is established and even if action comparison metric is known, it is still unclear to what segment of a video should both be applied? Action change detection segments a video into temporal boundaries, which no action change occurs in them, thus action representation is meaningful. This Partitioning, by counting the number of occurrences of each action class, can be useful for applications such as robotics, patient monitoring and athlete monitoring in sports centers.

As mentioned by Guo [2] "finding these temporal action boundaries is akin to scene cut detection in video, but in the space of actions". The detection of abrupt changes is a classical topic that has been studied in the past few decades [36]. It has been utilized in many areas, such as quality control, time series signal analysis, fault detection and monitoring, etc. [2]. In this paper, we propose an unsupervised action change detection method and introduce a new action representation based on the object silhouette sequence. We introduce a new representation of actions by displacing the axes and rotating the usual space-time volume. In this new volume, each frame consists of two axes x and t which

are correspond to height and width axes of that frame.

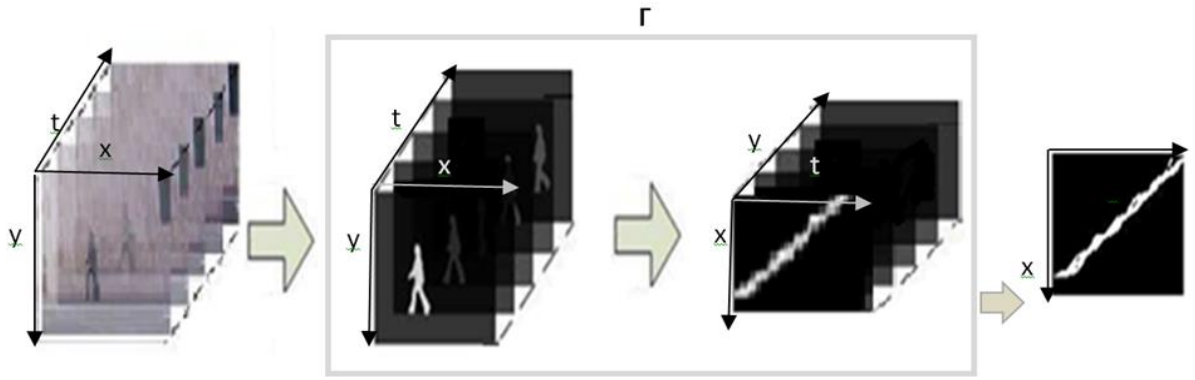

Fig. 2: An xt pattern extraction- Operator Γ shows rotating traditional space-time volume and displacing its axes.

In addition, y value specifies the frame number. It is show that this representation is not sensitive to silhouette noises, holes, and missing parts.

We created 105 artificial videos using the Weizmann dataset [37] to test the performance of the proposed method. There are 105 video segments with one action change, and 858 video segments without any action changes. Experimental results show 0% false negative error and 0.0023% false positive error. Therefore, the recall of our method is 100% and the precision of this method is 98.13%.

The rest of the paper is organized as follows. Next section reviews related works in action change detection. Proposed framework section describes the proposed action change detection method. Experimental results are reported in Results and Discussion section and concluding is presented in Conclusion section.

Related Works

In Guo et al. method [2], first, the silhouette sequence of video is broken into a sequence of overlapping N-frame-long action. For each point $s_0 = (x_0, y_0, t_0)^T$ into an N-frame action segment of silhouette tunnel a 13-dimensional feature vector is associated: three position features x_0, y_0, t_0, and ten shape features. The shape features are the distances between the point s_0 and the tunnel boundaries along the ten different spatio-temporal directions. Finally, a simplified representation for the shape of the silhouette tunnel will be obtained [38]. Next, a distance measure is defined to evaluate the similarity between any two consecutive action segments. The distribution of the pair-wise distance among in previous (T-1) segments is formulated using kernel density estimation. Lastly, a binary decision determines whether the segments S_1; S_2; . . . ; S_{T_1} and the segment S_T have continuous actions or not [2].

In order to measure the performance of this approach, they created 9 single person multi-action test video sequences. They used the segments of length for action comparison with a 4-frame overlap.

A judicious selection of these parameters is essential for the performance of their algorithm. It should be long enough to capture salient characteristics of an action, and to span only one single action. The precision of detected action boundaries depends on the length of action segments. Clearly, a large segment length increases the uncertainty of action boundary location. The total number of action segments is 597. 61 segments have action changes and 536 segments do not. The proposed action change detection method produces percent false negative error 1.64% and percent false positive error 0.19%.

Proposed Framework

Our framework for action change detection is based on the silhouette sequence of a deforming object. We produce silhouette sequence by using simple background subtraction techniques [39]. For example, three frames from a “two hands waving” action sequence and corresponding silhouettes from the Weizmann human action database [37] are shown in Fig.1. In this example, person is the foreground of the image and its corresponding pixels have a value equal to one (white).The pixels of background have a value equal to zero (black).

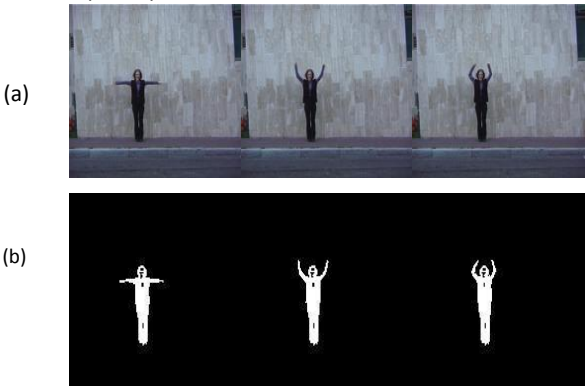

Fig. 1: Human action sequence. Three frames from (a) two hands waving action sequence and (b) corresponding silhouettes from the Weizmann human action database.

As shown in Fig. 2 instead of using the traditional space-time volume, a new representation is proposed by rotating this volume and displacing its axes. In the new volume, each frame point is localized by **x** and **t**, where **x** is the horizontal axes, like before, and **t** shows vertical axes now. In this volume, y values show frame number. The **xty** volume is a new representation of the traditional space-time volume.

To display an M * N frame, we use a two-dimensional array (matrix) with M rows and N columns .The value of each element of this array indicates the brightness of the frame at that point. This value can be zero or one in a binary image. The number of pixels with a value of one is the brightness of a frame .In the new volume, we select the highest brightness frame, as our suggested pattern. These patterns are similar together for each action. In these patterns, the silhouette sequence is condensed into a binary image. Therefore, it can represent motion sequence in a compact manner. In fact the silhouette sequence is condensed into a gray scale image. The quality of Silhouette tunnels is highly related to the correctness of background subtraction algorithms and the precise segmentation is very difficult to obtain in real world videos. Fig. 3 demonstrates an action sequence of "jumping-jack", which is chosen from the Weizmann human action database. As its corresponding xt pattern shows, this representation is not so sensitive to silhouette noise, holes, and missing parts.

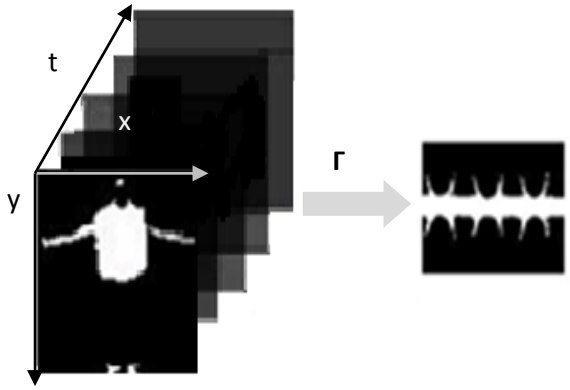

Fig. 3: The xt pattern is not sensitive to silhouette noise, holes, and missing parts.

Fig. 4 demonstrates several another examples of selected xt pasterns for different actions from the Weizmann human action database. The action sequences are "jumping- jack", "pjump" and "wave1". As their corresponding xt patterns show, these pasterns are completely different for different actions, so in proposed method, we work with this image data rather than silhouettes sequence.

If the action change occurs, the resulting pattern changes from a cross-sectional one. This cross- sectional shows the time of action change. Fig. 5, demonstrates several examples of the generated patterns that have action change (1: "wave-one- hand" to "wave-two-hands". 2: "Run" to "wave-two-hands". 3: "Run" to "jumping-jack". 4: "Run" to "walk").

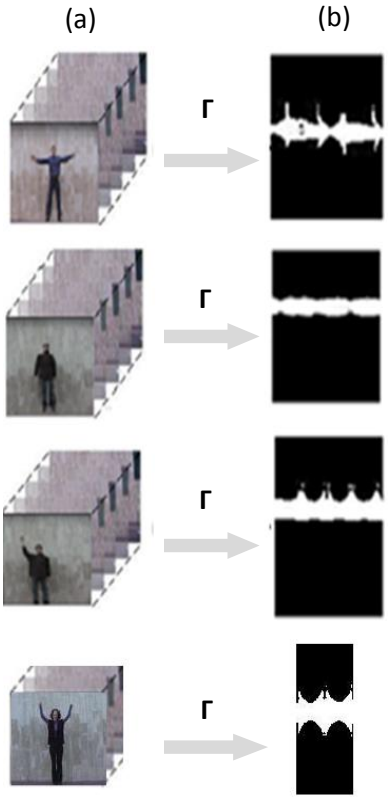

Fig. 4: Some examples of xt pattern obtained for different actions. (a) jumping-jack, pjump, wave1 and wave2 action sequences, from the Weizmann human action database. (b) Corresponding xt patterns.

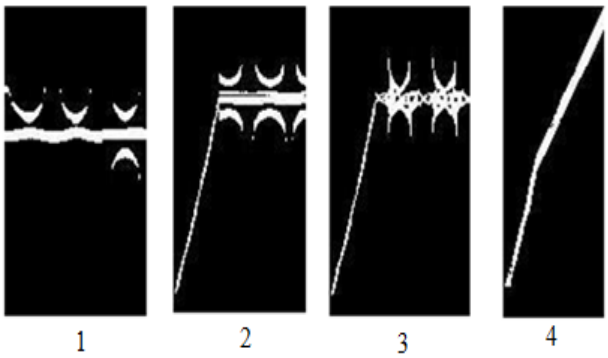

Fig. 5: Several examples of the generated patterns of videos that have action change. 1: "wave-one- hand" to "wave-two-hands". 2: "Run" to "wave-two-hands". 3: "Run" to "jumping-jack". 4: "Run" to "walk".

In order to find this cross-sectional, we use the histogram of oriented gradients (HOG) [17] feature vector. HOG describe appearance and shape by the distribution of intensity gradients or edge directions. We use the features = extractHOGFeatures(I) function in

MATLAB that extracts HOG features from an image I and returns the features in a 1-by-N vector. These features encode local shape information from regions within an image. The process of our method for action change detection can be summarized as follows:

The xt pattern is extracted from the xty volume. This pattern is splitted into several equal parts. The HOG feature vector is calculated for each part. The difference of the HOG feature of both consecutive parts (e.g. 1 and 2, 2 and 3 ...) is calculated by (1):

$$Dif = \sum_{i=1}^{SV} \left| HOG\_F_{P}[i] - HOG\_F_{P-1}[i] \right| \quad (1)$$

where, p is the part's number. HOG_F is the HOG feature vector and SV is the length of this vector. i is the index of each cell of HOG_F

All steps are repeated for the two parts that have the most difference value, i.e. they have the most difference in appearance

.If this maximum value is zero, the repeat is finished. At final, considering that each frame point is localized by x and t, the line that halves the last two parts represents the time that action change occurred.

The algorithm1 describes the proposed method for finding the part in which action has changed:

Algorithm1: Proposed algorithm for finding the part in which action has changed.

1. **Input**: A video sequence
2. **Output**: XT //the part in which action has changed.
3. XT= xt pattern of input video;
4. **While** XT
5. Split XT into several equal parts;
6. HOG_F:= ∅;
7. D: = ∅;
8. **for** P=1 to P ≤ NumberOf Parts; P++ do
9. F: = extractHOGFeatures (part of p);
10. Add F to HOG_F;
11. **if** (P>=2)
12. Dif= sum (abs (HOG_F (P) - HOG_F (P-1)));
13. Add Dif to D;
14. **end for**
15. MaxValue= maximum value of D Array;
16. **if** MaxValue > 0
17. XT:= part MaxIndex and part MaxIndex+1 of XT;
18. **else** break;
19. **end**

Results and Discussion

The proposed method implemented using MATLAB (R2015a) software in a Windows 8 operating system and a computer system with dual-core Intel Core-i7 processor and 8 GB RAM. In order to test the performance of the proposed method, we conducted a series of experiments based on ground-truth synthetic data as well as a time-continuous camera captured video. In the ground-truth experiment, we used the Weizmann dataset. We created 105 test video sequences. The database contains 90 low-resolution videos and silhouette sequences (180×144 pixels) that show 9 different people each performing 10 different actions, such as jumping, walking, running, skipping, etc. Some action examples are shown in Fig. 6.

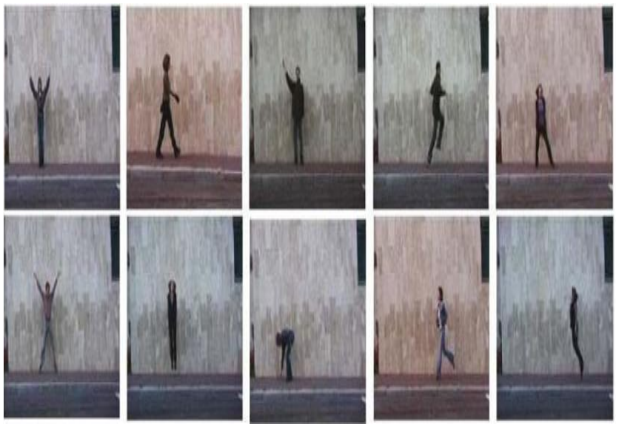

Fig. 6: Action examples from Weizmann dataset (wave2, walk, wave1, skip, side, jack, p-jump, bend, run and jump).

We use Precision and Recall for experimental evaluation of the proposed method. These measures can be calculated by (2) and (3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + FN} \times 100\,, \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + FP} \times 100, \quad (3)$$

where, TP is the true positive action change detection number. FP is the false positive action change detection number. False positive errors occur when a segment without any action change are classified as a segment with the action change. FN is the false negative action change detection number. False negative errors occur when a segment with the action change are classified as a segment without any action change. These functions are multiply to 100 to explain Precision and Recall in the percent measure.

To compare the results with the Guo's method [2], we assume generated video sequences as segments of length $N = 8$. There are 105 video segments with one

action change, and 858 video segments without any action changes. In Fig. 7, results of the proposed method on some sample video sequences are demonstrated.

The blue line indicates the detected time of action change by the proposed method.

As demonstrated in this figure, there is not any blue line in xt patterns of "Wave2", "pjump", "bend" and "jack", because action change doesn't occur.

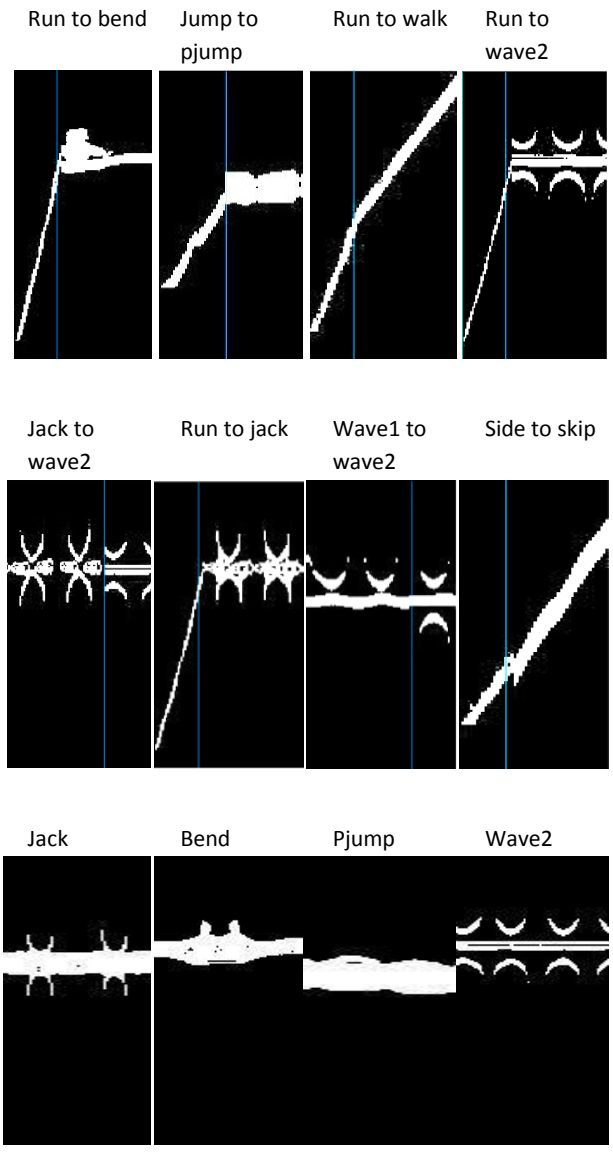

Fig 7: The results of the proposed method on some sample video sequence. The blue line indicates the action change detected time of the proposed method.

In the results of all video segments, there were two false positive errors as shown in Fig. 8.

In this figure there are blue lines in xt patterns of "jumping-jack" and "bend" but have not occurred any action change really.

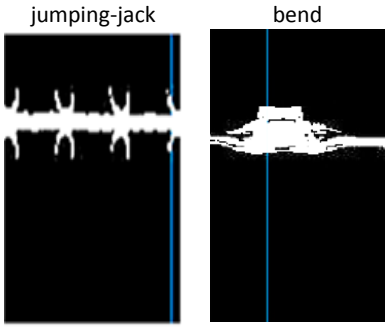

Fig. 8: False action change detection of the proposed method.

The proposed action change detection method produces 0% false negative error rate (PFN) and 0.0023% percent false positive error rate (PFP). Therefore, the recall of our method is 100% and precision of this method is 98.13%.

Table 1 shows efficacy of parameter P (number of parts) on precision of the proposed method. In general, the precision decreases with increasing P. Of course, there is a slight increase in P = 5 compared to P = 4. The precision of our method with P=4 is 77.78% and with P=5 is 78.4%. .The difference in HOG feature for very small parts is zero; therefore for P≥7 this algorithm hasn't good performance. Based on these results, this method has the best performance with P=3.

Table 1: Efficacy of parameter P on precision of the proposed method

P(number of parts)	precision of the proposed method
3	98.13%
4	77.78%
5	78.4%
6	66.68

Conclusion

The goal of action change detection is to partition a video into many sub-videos so that each of them contains only one single action. In this paper, we proposed a new approach to unsupervised action change detection in a video sequence. The proposed approach can automatically detect action changes without reference to labeled data. We introduced a new action representation method called xt pattern. The xt pattern is a selected frame of the xty volume. This volume is achieved by the proposed rotation of the space-time volume. A frame of this volume, with the most brightness, is our chosen pattern. These patterns are the same for each action and represents action sequence in a compact manner. In fact the silhouette sequence is condensed into a gray scale image. This representation is not so sensitive to silhouette noise, holes, and missing parts.

We created 105 test video sequences using the Weizmann dataset. Our experimental results indicate action boundary detection accuracy with PFN = 0% and PFP = 0.0023%. As mentioned, the proposed unsupervised approach can detect action changes with a high precision. Therefore, it can be useful in combination with an action recognition method for designing an integrated action recognition system.

In this research, we used the HOG feature. Given that the difference in HOG feature for very small parts is zero, proposition a more efficient algorithm to extract the feature of the xt pattern can increase the accuracy. In the future, it would be desirable to work on designing an integrated action recognition system using our action change detection method and an action recognition method.

Author Contributions

M. Fakhredanesh proposed the main idea of the innovation of the paper and designed road map of the research. M. Fakhredanesh and S. Roostaie designed the experiments and S. Roostaie implemented the experiments. M. Fakhredanesh carried out the data analysis.

Conflict of Interest

The author declares that there is no conflict of interest regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

Abbreviations

Γ	Rotating traditional space-time volume and displacing its axes.
Xt	The xt pattern is a selected frame of xty volume.
xty.	The xty volume is achieved by rotating the xyt space-time volume and displacing its axes
STEI	Stacked trajectory energy image
STV	Space-time volumes
FD	Fourier Descriptor
MEI	Motion energy image
MHI.	Motion history image
SWS	Statistical weighted segmentation
$BPNN$	Back-Propagation Neural Network
DNN	deep neural network
STIP	Spatio-temporal interest point
MRSR	Manifold regularized Sparse Representation
DBN	Dynamic Bayesian Network
HMM	Hidden Markov Model
CHMMs	Coupled hidden Markov models
ITBN	Interval temporal Bayesian network
DTW	Dynamic time warping
ANNs	Artificial neural networks
PMM	Part Movement Model
S_T	Tth action segment
PFN	False negative error rate
PFP	False positive error rate
P	Number of parts
HOG_F	The HOG feature vector
SV	The length of the HOG feature vector
i	The index of each cell of the HOG feature vector
TP	True positive
FP	False positive
FN	False negative

References

[1] K. Guo, P. Ishwar, J. Konrad, "Action recognition from video using feature covariance matrices," IEEE Transactions on Image Processing, 22(6): 2479-2494, 2013.

[2] K. Guo, Action recognition using log-covariance matrices of silhouette and optical-flow features. Boston University, 2012.

[3] S.-R. Ke, H. Thuc, Y.-J. Lee, J.-N. Hwang, J.-H. Yoo, and K.-H. Choi, "A review on video-based human activity recognition," 2(2): 88-131, 2013.

[4] Z. Weng and Y. J. J. o. E. I. Guan, "Trajectory-aware three-stream CNN for video action recognition," 28(2): 021004, 2018.

[5] H. Wang, A. Kläser, C. Schmid, C.-L. J. I. j. o. c. v. Liu, "Dense trajectories and motion boundary descriptors for action recognition," 103(1): 60-79, 2013.

[6] M. Ristivojevic, J. J. I. T. o. I. P. Konrad, "Space-time image sequence analysis: object tunnels and occlusion volumes," 15(2): 364-376, 2006.

[7] Y. Pritch, A. Rav-Acha, S. J. I. T. o. P. A. Peleg, M. Intelligence, "Nonchronological video synopsis and indexing," 11: 1971-1984, 2008.

[8] J. J. I. C. m. Konrad, "Videopsy: Dissecting visual data in space-time," 45(1): 34-42, 2007.

[9] M. Blank, L. Gorelick, E. Shechtman, M. Irani, R. Basri, "Actions as space-time shapes," in null, IEEE: 1395-1402, 2005.

[10] D. K. Vishwakarma, R. J. E. S. w. A. Kapoor, "Hybrid classifier based human activity recognition using the silhouette and cells," 42 (20): 6957-6965, 2015.

[11] N. Amraji, L. Mu, M. Milanova, "Shape–based human actions recognition in videos," in International Conference on Human-Computer Interaction, Springer: 539-546, 2011.

[12] A. F. Bobick, J. W. Davis, "The recognition of human movement using temporal templates," IEEE Transactions on pattern analysis, 23(3): 257-267, 2001.

[13] M. Sharif, Muhammad Attique Khan, Farooq Zahid, Jamal Hussain Shah, Tallha Akram., "Human action recognition: a framework of statistical weighted segmentation and rank correlation-based selection," Pattern Analysis and Applications): 281-294, 2020.

[14] C. C. A. Chen, J, "Recognizing human action from a far field of view," Proceedings of the 2009Workshop on Motion and Video Computing (WMVC)): 1–7, December 2009.

[15] S. Sehgal, "Human Activity Recognition Using BPNN Classifier on HOG Features," In Proceedings of the 2018 International Conference on Intelligent Circuits and Systems (ICICS),Phagwara, India): 286–289, 2018.

[16] M. A. Khan, Tallha Akram, Muhammad Sharif, Nazeer Muhammad, Muhammad Younus Javed, Syed Rameez Naqvi, "Improved strategy for human action recognition; experiencing a cascaded design," IET Image Processing: 818-829., 2019.

[17] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005, 1: IEEE): 886-893.

[18] Y. Zhu, W. Chen, G. J. I. Guo, V. Computing, "Evaluating spatiotemporal interest point features for depth-based action recognition," 32(8): 453-464, 2014.

[19] J. C. Niebles, H. Wang, L. J. I. j. o. c. v. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," 79(3): 299-318, 2008.

[20] L. Zhang, R. Khusainov, J. Chiverton, "Practical action recognition with manifold regularized sparse representation," in 29th British Machine Vision Conference: BMVC 2018, 2018: British Machine Vision Association, 2018.

[21] M. A. Khan, Kashif Javed, Sajid Ali Khan, Tanzila Saba, Usman Habib, Junaid Ali Khan, Aaqif Afzaal Abbasi. , "Human action recognition using fusion of multiview and deep features: an application to video surveillance," Multimedia Tools and Applications): 1-27, 2020.

[22] N. Hussain, Muhammad Attique Khan, Muhammad Sharif, Sajid Ali Khan, Abdulaziz A. Albesher, Tanzila Saba, Ammar Armaghan, "A deep neural network and classical features based scheme for objects recognition: an application for machine inspection," Multimed Tools Application, 2020,

[23] H. Arshad, Muhammad Attique Khan, Muhammad Irfan Sharif, Mussarat Yasmin, João Manuel RS Tavares, Yu-Dong Zhang, Suresh Chandra Satapathy, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," Expert Systems 2020.

[24] N. M. Oliver, B. Rosario, A. P. J. I. t. o. p. a. Pentland, m. intelligence, "A Bayesian computer vision system for modeling human interactions," 22, (8): 831-843, 2000.

[25] Y. Zhang et al., "Modeling temporal interactions with interval temporal bayesian networks for complex activity recognition," 35(10): 2468-2483, 2013.

[26] F. Negin, F. J. I. T. R. Bremond, "Human action recognition in videos: A survey," 2016.

[27] W. Ding, K. Liu, X. Fu, F. Cheng, "Profile HMMs for skeleton-based human action recognition," Signal Processing: Image Communication, 42: 109-119, 2016.

[28] Y. Zhou, A. J. P. R. L. Ming, "Human action recognition with skeleton induced discriminative approximate rigid part model," 83: 261-267, 2016.

[29] B. Saghafi, D. Rajan, W. J. P. A. Li, Applications, "Efficient 2D viewpoint combination for human action recognition," 19(2): 563-577, 2016.

[30] S. Das, M. Koperski, F. Bremond, G. Francesca, "A Fusion of Appearance based CNNs and Temporal evolution of Skeleton with LSTM for Daily Living Action Recognition," 2018.

[31] Z. Liu Z. Wang, "Action recognition with low observational latency via part movement model," 76(24): 26675-26693, 2017.

[32] S. Sempena, N. U. Maulidevi, P. R. Aryan, "Human action recognition using dynamic time warping," in Electrical Engineering and Informatics (ICEEI), 2011 International Conference on, 2011: IEEE: 1-5, 2011.

[33] S.-R. Ke, "Recognition of Human Actions based on 3D Pose Estimation via Monocular Video Sequences," 2015.

[34] D. C. Luvizon, H. Tabia, D. J. P. R. L. Picard, "Learning features combination for human action recognition from skeleton sequences," 99: 13-20, 2017.

[35] M. Hoai, Z.-Z. Lan, F. De la Torre, "Joint segmentation and classification of human actions in video," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE: 3265-3272, 2011.

[36] M. Basseville, I. V. Nikiforov, Detection of abrupt changes: theory and application. Prentice Hall Englewood Cliffs, 1993.

[37] L. Gorelick, M. Blank, E. Shechtman, M. Irani, R. J. I. t. o. p. a. Basri, and m. intelligence, "Actions as space-time shapes," 29(12): 2247-2253, 2007.

[38] K. Guo, P. Ishwar, J. Konrad, "Action recognition in video by covariance matching of silhouette tunnels," in XXII Brazilian Symposium on Computer Graphics and Image Processing, IEEE: 299-306, 2009.

[39] A. Elgammal, R. Duraiswami, D. Harwood, L. S. J. P. o. t. I. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," 90(7): 1151-1163, 2002. .

Biographies

Mohammad Fakhredanesh received his B.S., M.S. and PhD degree in computer science and Engineering from the Amirkabir University of Technology (Tehran Polytechnic), Iran, in 2005, 2007, and 2014 respectively. He is currently an assistant professor at the Malek Ashtar University of Technology. His research interests are the fields of computer vision, image processing, machine learning, and artificial intelligence.

Sahar Roostaie was born in April 1988. She received her B.S. degree in Computer Software Engineering, from Arak University in 2010, and M.Sc. degree in Artificial intelligence from MUT in 2017, respectively. Her research interests include the fields of machine vision, image processing and machine learning.

How to cite this paper:
M. Fakhredanesh, S. Roostaie, "Action change detection in video based on HOG," Journal of Electrical and Computer Engineering Innovations, 8(1): 135-144, 2020.

DOI: 10.22061/JECEI.2020.6949.351

URL: http://jecei.sru.ac.ir/article_1445.html