

PAPER • OPEN ACCESS

Human action recognition using a corners and blob detector with different classification methods

To cite this article: Zahraa Al-asady and Amel Al-amery 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **518** 052008

View the [article online](#) for updates and enhancements.

Human action recognition using a corners and blob detector with different classification methods

Zahraa Al-asady¹ and Amel Al-amery¹

¹Mustansiriyah University, Baghdad, Iraq.

Email: Zahraasalim29@gmail.com, de.amelhussein2017@uomustansiriyah.edu.iq

Abstract. Human Action Recognition trying to recognize the motion and movement of the person, the recognition of human movement is a very important field in many applications of security such as for smart surveillance and monitoring systems, also in another application such as computer interaction and entertainment. This paper tries to detect and identify the action of human but there is a big challenge in recognize similar actions for example walking and running. The dataset has been used is KTH dataset and the methods are harrier detector and blob detector to extract the corners from each frame. the second topic has been discusses in this paper is to examine between two different classification methods: supervised (SVM) and unsupervised (KNN) Machine Learning to find the most suitable way classification for this paper, we were find that the KNN classification is overcome the SVM Classification in this work . After the result has been extracted we find the using of KNN method is the most suitable for this topic. In future work we hope using more features to extract the corner to increase the efficiency of classification process.

1. Introduction

Recognize the human activity as the human do, is seems very interest project to do; it is the process of correctly recognize the action that has been performed by human [1]. It has been used in verity applications such as: surveillance issue to recognize the abnormal movement in many places such as Surveillance camera in roads, shopping, elderly houses, and healthcare systems [2, 3], it also used in other applications that depend on human-computer interaction such as games [4]. Last few 10 years the Concentration on human activity recognition increase spatially on the security field in the Context of “suspicious event” [5], “irregular behavior” [6], “uncommon behavior” [7], “unusual activity/ event/ behavior” [8]–[9], “abnormal behavior” [10]–[15]. “...activity analysis will be the most important area of future research in video surveillance” [1] this is so much true now days. In fact human action recognition is not new field, human action analyze has been interested from many. Centuries age but the first implementation by Gunnar Johansson [11] that used image sequence for human action analyze, in newest research in human action recognition by Forsyth et al. [13] focus on the recovery of human poses and motion from image sequences and also Ronald Poppe [4] that tried to label sequence of image by the name of action [3], Turaga et al. [12] focuses on the higher-level recognition of human activity.

The goal of all these researches is create robust system of human action recognition the main problem that he face is that the intra-classes variation [3] that’s mean there is very similar classes such as running and walking and within the same class we face problem such as the walking with different speed. The other challenges such as: the environment, lighting, and the same action can take from different views of camera



that will affect the classification among the classes specially the overlapped classes. So our research examines these challenges and show how it will effect on the classification of the action. This paper will discuss how to recognize the action using corner detector and store them in STIP descriptor then the last stage is machine learning algorithms using two type of classification methods: K-Nearest-Neighbor (KNN) method and Support vector machine (SVM), to test this work the dataset has been used is taken from KTH data set.

2. Material and methods

In this section we will disuse the data base has been used and the method for classification as shown in the flowchart bellow. We take the comparison between two classes and classify different classification method.

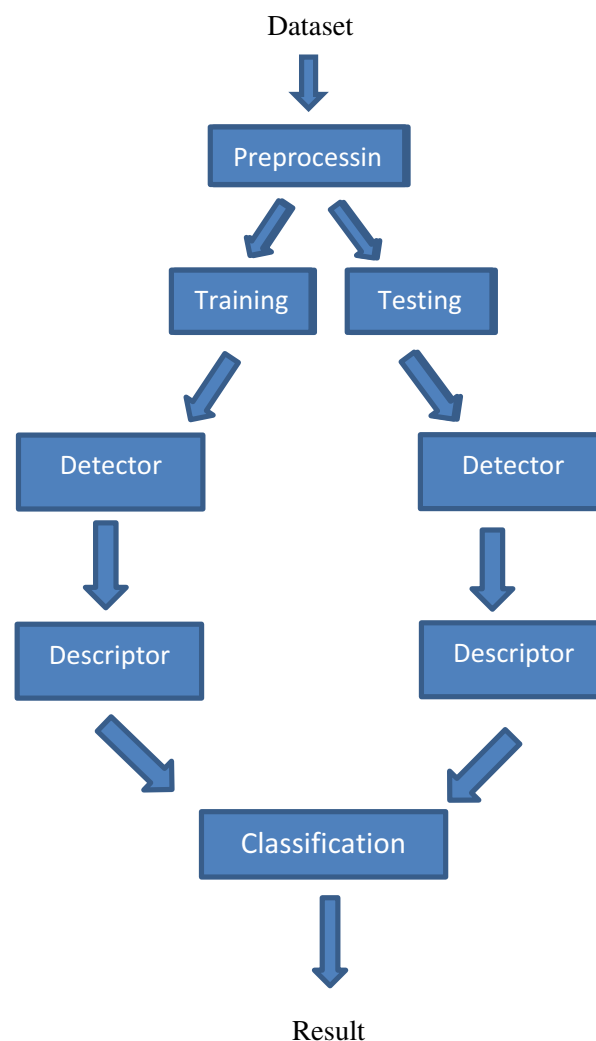


Figure 1. Flowchart that show the basic steps in human action recognition

2.1 Experimental database

The data base that has been used is the KTH human motion dataset [27], KTH is the most wide data base for human action it contain six classes for action (running, walking, boxing, jogging, hand waving, hand clipping) it contains almost 2391 sequence and also consider very strong data base because it is take 25 person and each person do the six action I four different cases (indoor, outdoor, zooming, different clothes). In this project subset of this dataset is taken to explain the challenge we deal with if we recognize human action Figure 2, so two similar classes and two different classes is taken which is (walking, running, boxing, and waving) the total number of video is 345.

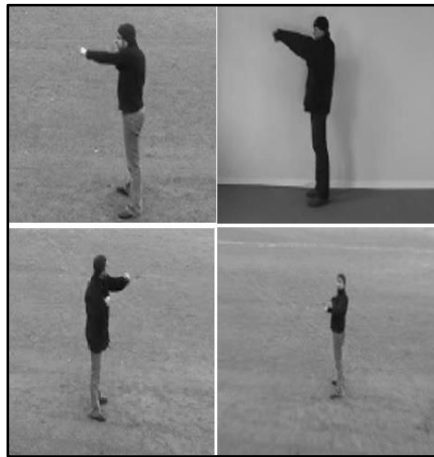


Figure 2. Shows boxing action of KTH data set in four scenarios outdoor, indoor, different clothes, and scaling.

2.2 Experimental framework

Human action recognition is very important field but it face a lot of difficulties and one of the main difficulties is the similarity between the action that the person can do such as walking and running so in this paper we study this difficulties and show the difference between classify the two similar actions and different action, based on the local feature to detect the edge to recognize the motion.

2.2.1 Preprocessing As mentioned before that's we depend on KTH video to recognize the motion , first step in preprocessing is dividing the video to frames to work on each frame Separated as shown in Figure 3.



Figure 3. Samples of frames of boxing video of the KTH data set.

2.2.2 Detectors As mention earlier that the aim of this work is to detect and recognize the human action, and to do that we focused on extracting the corner therefore two type of detect has been used: Harris detector and Hessian detector, Harris detector was introduced by Laptev and Lindeberg [24], where harrier detector depends on compute for each point in the frame a spatio-temporal second-moment matrix.

$$\mu(\cdot; \sigma, \tau) = g(\cdot; \sigma, \tau) * (\nabla L(\cdot; \sigma, \tau) (\nabla L(\cdot; \sigma, \tau))^T) \quad (1)$$

Where σ, τ are independent spatial and temporal scale values, g is Gaussian smoothing function, and ∇L is space-time gradients, the last stage is local maxima of H that determine the final location of interest point from harrier detector

$$H = \det(\mu) - k \text{trace}_3(\mu), H > 0 \quad (2)$$

Hessian detector was introduced by [25], this detector is used for extract the blob from the images The position and scale of the interest points are simultaneously localized without any iterative procedure.

$$\det(H) = I_{xx}I_{yy} - I_{xy}^2 \quad (3)$$

Where I_{xx} , I_{xy} , and I_{yy} are second-order image derivatives computed using Gaussian function of standard deviation σ .

2.2.3 Descriptor

To complete the classification methods we need features to be extracted, stored, and ready to be used by the classifier, each object in video or frame can extract some features from it, this is done by the detector and the second important step is represent these features in numerical vector Scale-Invariant Feature Transform (STIF) descriptor is used, where it convert each feature into 128-dimensional vector.

2.2.4 Classification methods and result

The frame is represent as feature that has been extracted then moving to next step which is classification stage, classification is done by steps: the first step is classifier learning and that's what refer to as "training process", the feature has been extracted from the training set enter to classifier for learning process to distinguish different groups, the training process should be robust and solid that why the training set is larger than testing set, second step in classification is test the classifier based on the similarity between the testing feature and the training feature. This paper takes two type of classification method and compares the result between them 2.2.4.1 And 2.2.4.2 these sections explain the classification method: KNN and SVM.

2.2.4.1 K-Nearest-Neighbor Classifica

K-nearest neighbor algorithm [26] it is method that classifies the object based on closest point that exists in feature space (training set). Its consider to be the easiest learning algorithm[26], as mention before the classification method basically relies on training data that enter to the KNN algorithm which is: feature vectors and label of the training frames of each video, testing step takes the feature vectors of the test video and label them base on its k nearest neighbors. Typically the classification of the object depends on the labels of its k nearest neighbors by majority vote. If k=1, the object is simply classified as the class of the object nearest to it. When there are only two classes, k must be an odd integer. However, there can still be ties when k is an odd integer when performing multiclass classification. After we convert each image to a vector of fixed-length with real numbers, we used the most common distance function for KNN which is Euclidean.

$$\begin{aligned} d(x, y) &= \|x - y\| = \sqrt{(x - y) \cdot (x - y)} \\ &= (\sum_{i=1}^m ((x_i - y_i)^2))^{1/2} \end{aligned} \quad (4)$$

Where x and y are histogram in $X=R^m$. Advantage of KNN is works well with multiple model classes Couse the resolution is depends on a little neighborhood of corresponding class [26]. Therefore, if the target group is multi-modal, the KNN algorithm leads to good result and accuracy also when we have huge data its works better and consume run time then SVM. While the main disadvantage of KNN algorithm its uses when finds the similarity all the features evenly [19] this technique leads to miss classification, especially when small subset of features which is really help in classification is feature space.

2.2.4.2 Support Vector Machine Classification

Support Vector Machine (SVM) classifier [26] different planes in space have used for dividing the data points [14]. This model is a represent of the data in space as dots, mapped in such a way where data of the different categories are separated by a separating plane that increases the border between different groups [12]. This is due to the reality of, if the dividing plane has the largest distance to the nearest training points of one of the classes, it decrease the generalization error of classifier. The test points are represent into that same feature space and predicted to belong to which category based on which side of feature space falls in to as shown in Figure 4. Given a training set of instance-label pairs (x_i, y_i) , $i=1..l$ where x_i and y_i are histograms from training images [26], the support vector machines can be found as the solution of the following optimization problem:

$$\begin{aligned} \min_{w, b, \zeta} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \zeta_i \\ & y_i (w^T \phi(x_i) + b) \geq 1 - \zeta_i, \quad \zeta \geq 0. \end{aligned} \quad (5)$$

The linear kernel function and the radial basis function can be represented as,

$$K(x_i, x_j) = x_i^T x_j \quad (6)$$

$$K(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}, \gamma > 0 \quad (7)$$

SVM classification contain many advantages the main advantage is that SVM do well when the datasets have many attributes, even if few available class only exist in space in the training process [12]. Also there are disadvantages in SVM method such as limitations of runtime speed and size in training and testing process other challenge in SVM method is kernel function parameters selection.

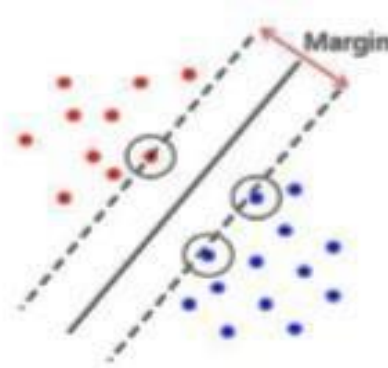


Figure 4. SVM Classification. In multidimensional space, support vector machines find the hyperplane that maximizes the margin between two different classes.

3. Experimental results

In this section different classification method has been used on the data of descriptor/ detectors, we should mention that the challenge in recognize the human action is the similarity between the actions such as (running and walking) as shown in Figure 5.

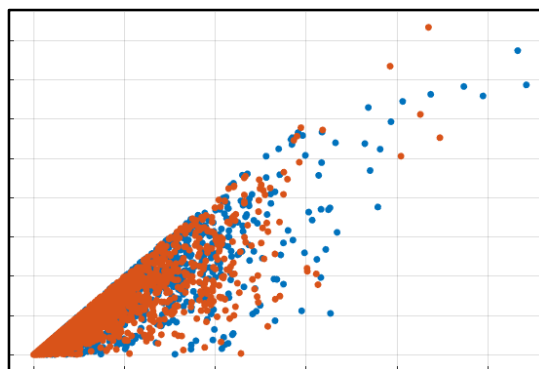


Figure 5. Plot the data of two similar classes (walking and running).

The idea here to take different human actions to recognize by the classifier and at the same time explain the challenge of classification of human movement so the classes take in this study is four two similar (walking and running) and different classes (waving and boxing). The total number of feature almost 8300 frame of

video where 1660 image for testing and 6640 image for training process. And the proposed method tries to detect the corner and blob in each frame and then using local maxima to extract the strongest point only of each frame these feature how inter to classifier showing the result below A comparison between the classification methods is done and the results are shown in table (1), also confusion matrix illustrative between two different and similar classes in Figure 6.

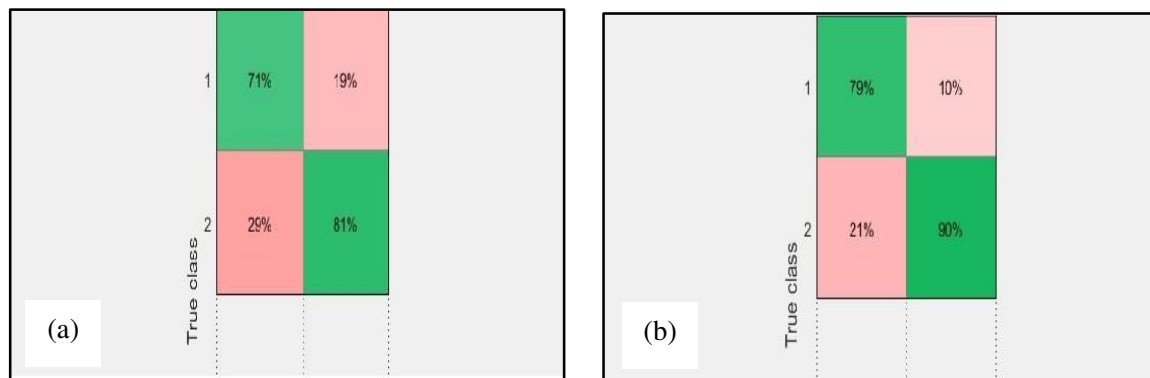


Figure 6. Confusion matrix of KNN classifier, (a) the result of classification of similar action (running and walking), (b) the result of classification of different classes (boxing and waving).

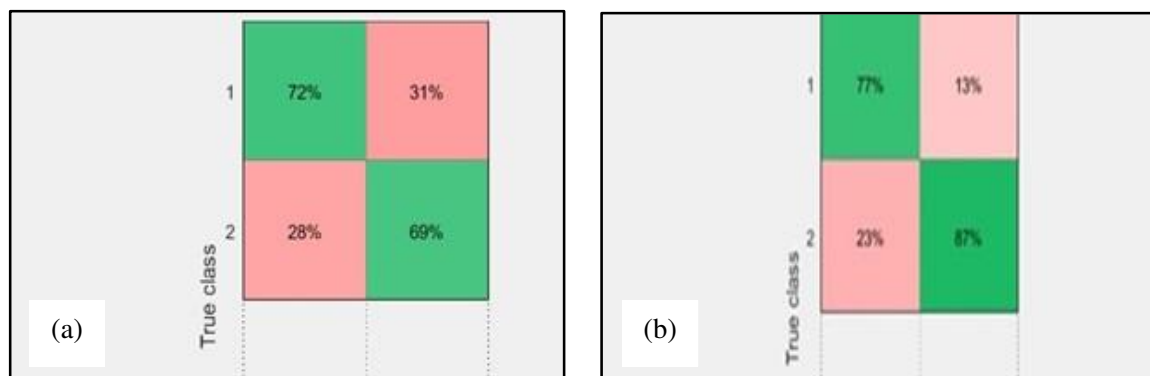


Figure 7. Confusion matrix of SVM classifier, (a) the result of classification of similar action (running and walking), (b) the result of classification of different classes (boxing and waving).

Table 1. Show the comparison between classification methods with different human action

	Different classes	Similar classes
KNN	87	75
SVM	81	71

4. Conclusion

In this work, we detect and classify human action by extract the corner and blob features with suitable learning algorithm, and we could observe that KNN is overcome SVM also the KNN performance is faster than KNN and the big challenge we face here is classify the similar classes, where the different in accuracy

of classification of KNN between similar and non-similar almost 12% and in SVM 10%. We know to improve the classification process it is depend on characteristic of the data that enter to it so there is no classification method that work perfectly with the problems, as mention there is a lot of challenge in this subject and we still try increase the performance. In future work we would use different descriptor and more robust classification method.

Reference

- [1] Qluwatoyin popoola and kejun wang 2012 *Video based abnormal human behavior recognition* IEEE.
- [2] Jesper Birksø Bækda 2016 *Human Action Recognition using Bag of Features* IEEE.
- [3] Ruoyun Gao 2009 *Dynamic Feature Description in Human Action Recognition* Dynamic Feature Description in Human Action Recognition Leiden Institute of Advanced Computer Science.
- [4] Ronald Poppe and Mannes Poel 2008 *Discriminative human action recognition using pairwise CSP classifiers*, in *(FGR'08)* September 2008.
- [5] Zhang Y and Liu Z 2007 *Irregular behavior recognition based on treading track* in *Proc.Int.Conf.WaveletAnal.PatternRecog.*
- [6] Willem A, Madasu V, Boles W and Yarlagadda P 2008 Detecting uncommon trajectorie in *Proc. Digital Image Computing: Tech. and Applications*.
- [7] Zhong H, Shi J and Visontai M, "Detectingunusualactivityinvideo," in *Proc. 2004 IEEE Comput. Vis. Pattern Recog.*, 2008.
- [8] Hara K, Omori T and Ueno R 2002 Detection of unusual human behavior in intelligent house," in *Proc. 12th IEEE Workshop Neural Netw. Signal Process.*, 2002.
- [9] Y. Hung, C. Chiang, S. J. Hsu, and C. Chan, "Abnormality detection for improving elder's daily life independent," in *Proc. 8th Int. Conf. Smart Homes Health Telematics*, Jun.22–24, 2010.
- [10] Lee C K, Ho M F, Wen W S and Huang C L 2006 Abnormal event detection in video using N-cut clustering *Proc. Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*
- [11] Haralick R 1983 *Ridges and Valleys on Digital Images* Computer Vision, Graphics, and Image Processing.
- [12] Turaga P K, Chellappa R, Subrahmanian V S and Udrea O 2008 Machine recognition of human activities: a survey. *Proc. IEEE Trans. Circuits Syst. Video Technol.*
- [13] Forsyth A and Fleck M M 1997 *Body plans* In *IEEE Conference on Computer Vision and Pattern Recognition* 678–683 Puerto Rico.
- [14] Lei Wang , Yangyang XU, Jun Cheng, Haiying XIA and Jianqin Yin 2018 *Human action recognition by learning spatio-temporal Feature With Deep Neural Networks* IEEE.
- [15] Satyabrata Maity, Debotosh Bhattacharjee and Amlan Chakrabarti 2015 *A Novel Approach for Human Action Recognition from Silhouette Images* Scopus.
- [16] Yap T, Jiang X and Kot A 2010 *C Two-dimensional polar harmonic transforms for invariant image representation* IEEE Trans Pattern Anal.
- [17] Kim B, Yoo H and Sohn K 2013 *Exact order based feature descriptor for illumination robust image matching* Pattern Recogn.
- [18] Lazebnik S, Schmid C and Ponce J 2006 Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories *Computer Vision and Pattern Recognition* IEEE Computer Society Conference.
- [19] Niebles J C, Wang H and Fei-fei Li 2008 *International Journal of Computer.*
- [20] Laptev I, Marszalek M, Schmid C and Rozenfeld B 2008 Learning realistic human actions from movies *IEEE Conference on Computer Vision and Pattern Recognition*.
- [21] Jose M Chaquet A, Enrique J, Carmona A and Antonio Fernández-Caballero 2013 *A survey of video datasets for human action and activity recognition*" Scopus.
- [22] Laptev I and Lindeberg T 2003 Space-time interest points *ICCV*.

- [23] Willems G, Tuytelaars T and Van Gool L 2008 An efficient dense and scale-invariant spatio-temporal interest point detector *ECCV*.
- [24] Bremner D, Demaine E, Erickson J, Iacono J, Langerman S, Morin P and Toussaint G 2005 Outputsensitive algorithms for computing nearestneighbor decision boundaries *Discrete and Computational Geometry*.
- [25] Hassaballah M, Aly Amin Abdelmgeid and Hammam A 2016 *Image Features Detection, Description and Matching* Springer.
- [26] Schuldt Laptev and Caputo 2005 *Proc. ICPR' 4* Cambridge UK.