

Image Based Human Action Recognition

Deepshikha¹, Brijesh Pandey², F. Masooma Nigar³P.G. Student, Department of Computer Engineering, GITM, Lucknow, Faizabaad road, Uttar pradesh, India¹Associate Professor, Department of Computer Engineering, GITM, Lucknow, Faizabaad road, Uttar pradesh, India²Associate Professor, Department of Computer Engineering, GITM, Lucknow, Faizabaad road, Uttar pradesh, India³

ABSTRACT: Detecting human action is a part of computer vision is a challenging problem. Many machine learning methodologies have been adapted to implement this problem. Most of the research has been towards human action recognition via moving images or video frames. Detecting human action finds applications in robotic vision, surveillance system etc.

The main concept behind action detection is to find out the body shape, pose, angle etc. features. Once these features are obtained any machine learning algorithm can be used to classify the objects in action. But this problem seems to be difficult in still images where there are challenges posed by background of the image. In still images there are no displacements in the body pose. Hence finding action in still images is a challenge. In this thesis work, I propose a method of still images based action recognition system through support vector machines. Support vector machines are linear classifier. Thus it is again a challenge to modify the model of linear svm and train it to classify multiple classes' data set.

KEYWORDS: SVM, Machine learning, Kernel..

I. INTRODUCTION

A visual surveillance system is considered as an important part of human action recognition. Recognizing human action is an integral part of Computer vision application. Action recognition involves analysis of ongoing event from the given data. Over the years researches have successfully implemented the action recognition via video sequences. Whenever the some unsuitable circumstances are present like patrolling system process at boarder, nuclear reactors etc. then the Visual surveillance systems play a very crucial. Whenever there need a surveillance system for the some business like parking lot monitoring, surveillance in shopping complex then there always demand for automatic surveillance systems. To collect the data and complete the process by manpower is not so easy, manpower system done the process by monitor the data collected from various cameras continuously. Completely understanding about the human actions and building a higher level knowledge of the events is tough. Computer vision system completed the all process in a very easy way through the occurrences of scene.

One can find that many action categories can be sketch unambiguously in single images without motion or without the need of video signal, human perception is based on the theory of video signal. Figure 1, depicts some images from which one can infer that one image is having a human riding a bike and other playing music. Hence action can be represented from any still image. This motivates the development of computational algorithms for automated action analysis and recognition in still images. We all know that internet has many images of humans in action. We can use them to analyze human action in those images.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016



Figure 1.1: riding a horse, running, images depicting human in action. [8]

Human action recognition is a great challenge for computer vision. There have been many researches over the years. Most of the efforts have been towards action recognition through moving images. Some efforts have been made for action recognition via still images but still have not been up to the mark. Hence in this work we propose still images based human action recognition via Support vector machine and compare the performance of the two variants of support vector machine kernels, the poly kernel and RBF kernel.

II. LITERATURE REVIEW

There have been a lot of researches done on the topic of Human Action/Activity Recognition by various scientists all over the world. Various methods have been used to identify the action posed, most of the or more or less every paper focused on the image acquisition from a video or image sequence and then the templates were created from the video stream and the pose to be determined was extracted from the frames and then the matched frame was then again converted to the video stream and classification was done, on the base of which the result was produced.

The first attempts began with the concept of background subtraction which was used to separate background and foreground image which helped in motion segmentations. (Heikkila & Silven, 1999) [15] Developed algorithm for background subtraction. This posed many drawbacks .thus a year later(Collins.J.J.,1997).) [12] Attempts were made and new algorithm for temporal differencing was developed. It is used to separate video frames in a constant time.

Many efforts were on but most of the research work was video based action recognition. (Aaron F. Bobick&Davis,2001)[2] Brought in a new concept of state based method using Hidden Markov model for learning visual behaviour from image sequences.

This motivated towards human action recognition using still images. (Gupta, A. Kembhavi, & Davis,2009), (Ikizler&Duygulu,2008), [1,13] Proposed still images base action recognition for specific activity of sports action recognition. Body pose was used as a cue for action recognition in these methods.

(L.J. Li& L.Fei-Fei,2007)[14] Described the use of scene types in categorizations of actions which can be used for action recognition. The following proposed work extends these works by incorporating SVM and SURF points for human action recognition.

The paper (Siley Oumar Ba) [4] titled, states the use of Visual Focus of Attention (VFOA) technology for the use of Human Activity Recognition. VFOA of an individual provides an insight about what is the enthusiasm of an individual, who is the focus of the individual's discourse, and who the individual is listening to. The fundamental concentrate in this paper is to study about individuals' VFOA utilizing machine vision strategies. To gauge this from a machine vision perspective, following the individual's look is required. As, following the look of eye is unthinkable on low or mid determination pictures, we utilize head introduction as a route around for the look course. In this paper we examine about following the introduction of individuals' head and as a second step the VFOA of the individual is perceived from their head introduction. Once the introduction of head posture is accessible, VFOA recognition could be possible. Two situations have been utilized here to study the VFOA: A gathering room environment and an open air environment.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

In the gathering room environment, individuals are static. Individuals' VFOAs were concentrated on relying upon their areas in the gathering room. The head stances are utilized as perceptions and potential VFOA focuses as concealed states in a Gaussian mixture model (GMM) or a shrouded Markov model (HMM) structure. In the open air environment, individuals are moving and there is a solitary VFOA target. The issue in this study is to track different individuals passing and assessment whether they were centering the notice. The VFOA is displayed as a GMM having as perceptions individuals' head area and posture.

The paper (Arasanathan Thayanathan) [5] states about using the templates created from the images acquired by a camera using hand gestures as the activator to extract the poses and estimate the result. The paper focuses on the problem of tracking three dimensional human hand motions from singular view. This paper focuses on solving the tracking part in a unified framework. They then use tracking the hand movement for the pattern extraction for solving this problem. Template matching forms the basic building block of the proposed systems in this work.

The primary methodology assembles a progressive Bayesian following framework in a generative structure, which approximates the back dissemination at different resolutions utilizing layouts. The focal point of this representation is that districts with low likelihood mass can be quickly disposed of in a various leveled inquiry, and the dispersion can be approximated to self-assertive accuracy. The adequacy of the strategy is exhibited by utilizing it for following 3d verbalized and worldwide movement before jumbled foundation. The second approach proposes a discriminative system where an inadequate representation of the formats is found out. With a specific end goal to address the issue of posture equivocality, a one-to-numerous mapping from gimmick to state space is found out utilizing a situated of pertinence vector machines. The picture gimmicks are acquired by vigorous layout matching, and the pertinence vector machines select a meager set of these layouts. The relapse system is connected to the stance estimation issue from a solitary information outline, and is implanted inside a probabilistic following structure to incorporate transient data.

The paper (6 Benjamin John Sapp) [6] states about the use of images for extracting the pose out of them and declaring what pose is enacted in the image. Estimating the Human pose from 2D images has always remained and is one of the most challenging aspects of computer science and computationally demanding problems in computer vision. In this paper, we propose a novel method which allows the user for efficient inference in richer models with data-dependent interactions. The paper tried to focus on the possibilities and capabilities of the computer vision system to extract the pose pattern from the image given but was not able to, hence the image sequence model has been used to capture the frames from the sequence and then extract the pattern from those frames extracted.

III. METHODOLOGY

The proposed system is a still image based action recognition system. It is represented as IHARS. It is built using the technologies of Computer Vision, image processing toolbox of MATLAB and Artificial Intelligence. There exists many methods for the proposed activity recognition, but all of them are based on the stream of images or multiple video sequences to observe and infer out the activity performed, but what is being tried to achieve from the proposed system is an efficient human action recognition system which can be incorporated in the development of visual system of any robot. It is not an easy task to accomplish, as in the case of motion picture or a video there are multiple sequences available to judge what pose is actually performed, but in the case of a still image it is all dependent on the systems capabilities to perform the computation and observe the pose which is most likely performed and then produce the desired result.

The proposed image based action recognition system (IHARS) follows the following steps to tell us about the action of the object from a still image.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

Algorithm 1: proposed methodology of IHARS

- Step 1 download images of human in action from internet and store them as training data set. Label them as a class label. Store the details in .CSV file format. Repeat the same step for test images
- Step 2 : For differentiating human in action with background part of the image make a rectangular box around the man in action using SURF points object detection methods.
- Step 3: the new database with human in action represented in rectangular box is selected and image histograms are obtained such that we have two histograms one for human in action as object histogram and the other representing the background of the image as visual histogram.
- Step 4: histograms are normalized for L1 norm.
- Step 5: Multiclass SVM is trained for L1 Norm histogram in one vs. All manners through histogram intersection kernel. Use poly kernel and RBF kernel for classification over same data.
- Step 6: Now the training of the SVM is done for each action input image in 1vs all manner. There are two outcomes positive and negative. Positive represents human action negative other parts of the image other than human action.
- Step 7: Testing is performed with action detection. Display the result of matching performed over test images.

Pre-processing of the data set

The dataset consists of images of human performing certain kind of action. The dataset is broken down into two different sets; one as training dataset and other as testing dataset.

The images are downloaded from the internet, where in human is performing certain kind of action like walking, jumping, running, riding horse etc.

Six different actions have been selected they are

1. Human interacting with computer
2. Human jumping
3. Reading
4. Riding a horse
5. Running
6. Walking

Images based on the above actions are downloaded in JPEG format and stored as testing set. Each image is assigned an action label in numeric form like 1, 2, and 3, so on where in 1 specifies that the human in action is interacting with computer, 2 specify human is jumping and so on.

Similar kinds of images are downloaded to form the testing set.

Images of poor quality are deleted. Each image is annotated with rectangular boxes for identification of the human using the SURF based object segmentation. For each rectangular box an attribute action is used to define the action being performed.

Strategies of incorporating the human in the rectangular box:

1. Human is considered as object: each image is chosen where in a human is performing certain action, the image is resized till 500 pixels. The human in action occupies majority of the image and background of the image is not neglected.
 2. Image histogram: it is used to represent action and background of the image.
 3. Image segmentation: the process of image segmentation is carried out to differentiate the foreground and the background part of the image. In the foreground we have human performing action. Hence we have an object histogram of the image which represents the histogram of the foreground. The other part of the segmentation involves the background part and it is represented by visual histogram.
- Each image is resized so that it can be easily classified by the multi-class SVM.

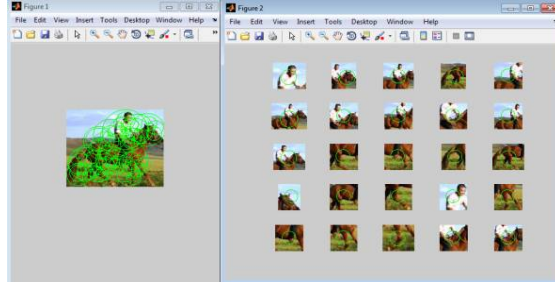
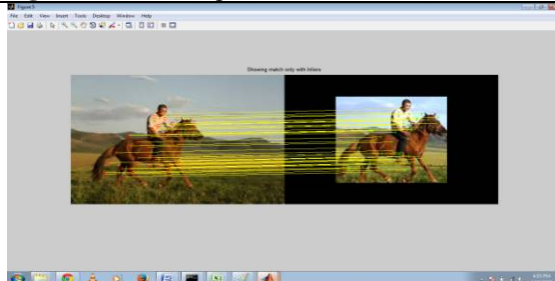
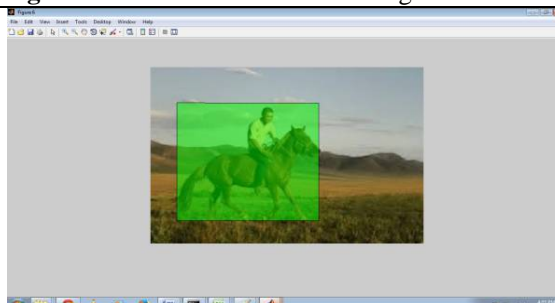
International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

IV. EXPERIMENTATION & RESULT

The tests have been carried out on a Intel core i5 processor with 4GB RAM running a 32 bit windows 7 operating system with MATLAB R2013b with supported MEX files for Visual C++.

Figure	Description
 <p>Figure 1: Extracting Surf Point and Blob</p>	<p>Speeded Up Robust Feature (SURF) point and blob provide complementary information about regions in a digital image.</p>
 <p>Figure 2: The SURF Points matching.</p>	<p>Speeded Up Robust Feature (SURF) use for object recognition and classification.</p>
 <p>Figure3:The rectangular bounded image of human in action.</p>	<p>The objective is to detection human in action and make rectangular box.</p>

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

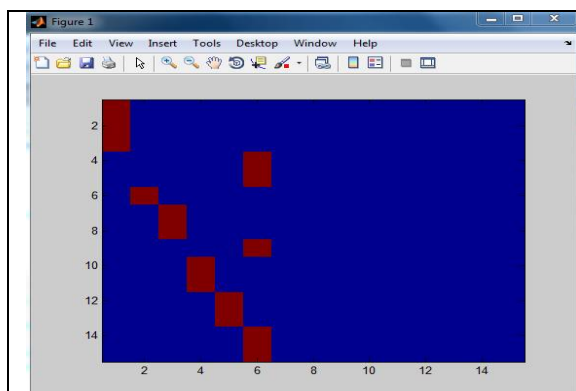


Figure 4: Confusion Matrix

Classified as (a, b, c, d) computer, jump, Read, Riding, Running, Walking using matrix. Thus with RBF Kernel SVM the output obtained is 80% and that with poly kernel the output is 16.667% correctly classified.

Result:

Action	Number of matches
Interacting with computer	Out of 3 all matched perfectly
Interacting with computer	Out of 3 all matched perfectly
Jumping	Out of 3, 1 matched correctly, 2 wrong matches to walking class
Reading	Out of 2 images , all matched correctly
Riding horse	Out of 3 , 2 matched perfectly and 1 matched to walking class
Running	Out of 2 images , all matched correctly
Walking	Out of 2 images , all matched correctly

In total out of 15 images 3 images were wrongly matched to a different class. This gives the accuracy for 80 %. Hence the proposed method is far better classifier method than the existing RBF based kernel method.

V. CONCLUSION

This research work analyzed the performance of the Multi-class SVM based classifier used on the SURF based blob images of humans in action. The main task of action recognition in still images was challenging. To test the proposed system a dataset of humans in action was created with 6 actions. Each action was assigned a label.

The work successfully demonstrated the recognition of human action from given image based on the training dataset. It seems to be successfully helpful in recognizing human in action from still images. The overall efficiency has been described in table 1 where classification map based on the action label for different classes of actions is mentioned. The efficiency of the proposed system can be further enhanced by the choices of kernels. It can be improved by treating and matching the foreground and background separately using two separate kernels. This holds for all classes except "Running" and "Walking" where using background slightly increases the confusion with the other action classes. SURF based object detection is time consuming but it gives us many positive results any it can be incorporated with other techniques to improve the overall efficiency.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

REFERENCES

- [1]. Gupta, A. Kembhavi, and L.S. Davis(2009). Observing human-object interactions: Using spatial and functional compatibility for recognition. IEEE PAMI, 31(10):1775–1789, 2009.
- [2]. Aaron F. Bobick and James W. Davis(2001).The recognition of human movement us-ing temporal templates. IEEE Trans. Pattern Analysis Machine Intelligence, PAMI, 23(3):257–267, Mar 2001.
- [3]. IvanovY.A., Bobick ,A.F. (2000), Recognition of visual activities and interactions by stochastic parsing, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (8):852 – 872..
- [4]. SileyOumar Ba(2006),Joint Head Tracking and Pose Estimation for Visual Focus of Attention Recognition,EcolePolytechniqueF´ed´erale de Lausanne
- [5]. ArasanathanThayananthan(2004) , Template-based Pose Estimation and Tracking of 3D Hand Motion, University of Cambridge
- [6]. Benjamin John Sapp(2007) , Efficient Human Pose Estimation with Image-dependent Interactions, University of Pennsylvania
- [7]. Hemali S. Mojdra and Borisagar,V.H.(2012), A Literature Survey on Human Activity Recognition by means of Hidden Markov Model, Universal Conference in Recent Trends in Information Technology and Computer Science (ICRTITCS – 2012)
- [8]. Collins.J.J.(1997). Robot pursuit and evasion using temporal difference learning, submission at GEG University of Limerick Foundation(ULF)
- [9]. N.Ikizler and P.Duygulu(2008), Recognizing actions from still images,inproc.ICPR.
- [10]. Li,L.L. and Fei-Fei,(2007)., what,where and who? ,classifying events by scene and object recognition ,ICCV
- [11]. J.Heikkila and O.silven(1999), Background subtraction techniques,second IEEE journal,(1999).
- [12]. PolanaR., NelsonR., Low level recognition of human motion, Workshop on Non-Rigid Motion. 77 – 82, 1994.
- [13]. R. Rosales, Recognition of human action based on moment based features, Technical Report BU 98-020, Boston University, Computer Science, 1998.
- [14]. R. VenkateshBabu, B. Anantharaman, K.R. Ramakrishnan, S.H. Srinivasan, Compressed domain action lassification using HMM, Pattern Recognition Letters 23 (10):1203 – 1213, 2002.
- [15]. S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice Hall, 1998.