ISSN 2395-1621

Human Interaction Recognition Using Patch-Aware Models

[#1]Shweta Upare, [#2]Sneha Kompelli, [#3]Mayuri Mokashe, [#4]Prof. S .K. Patil
[#5]Dr. Y. S. Angal

[1]upareshweta0084@gmail.com

[#123]Student, Department of E&TC
[#4]Asst. Professor, E&TC Department
[#5]Head of ENTC Department

JSPM's BSIOTR, Pune, India.

ABSTRACT

This paper describes the human interaction recognition with close physical contact from video frame. Because of frequent occlusions and feature to person assignments it is difficult to recognize the human interacting activities. Due to this recognition performance will degrades. We propose a novel framework for recognizing the close human interaction. Our model collaborate a set of hidden variables with spatiotemporal patches and accurately conclude their states, which specify the person that belongs to patches. Thus our model overcomes the problem of uncertainty in feature assignments. Moreover, we cover the earlier for the patches to deal with frequent occlusions during interaction. Our model recognizes close human interaction using discriminative supporting regions.

Index Terms-action recognition, human interaction, frequent occlusions, and spatiotemporal patches.

ARTICLE INFO

Article History

Received: 24th March 2017

Received in revised form :

24th March 2017

Accepted: 26th March 2017

Published online :

8th April 2017

I. INTRODUCTION

Human activity recognition aims at building robust and efficient computer vision algorithm and systems which can automatically recognize specific human activities from a sequence of video frames. High-level understanding of human activity is essential for various applications, including surveillance systems and human computer interactions [1]. In particular, a human activity recognition system may enable the detection of abnormal activities as opposed to the normal activity of persons using public places like airports and subway stations. Automated human activity recognition may be useful for real-time monitoring of the elderly people, patients, or babies. Human action recognition has been of great interest for the computer vision community for many decades due to its practical importance, such as video analysis and visual surveillance[2].

A majority of action recognition target on analyzing the action then fully observing the entire video. However, in very real-world scenarios (e.g. vehicle accident and criminal activity), brilliant systems do not have the comfort of waiting for the entire video before having to answer to the action contained in it. For example, being able to suppose a serious driving stage before it arise; opposed to recognizing it thereafter. Awkwardly, most of the present action recognition suggestions are improper for such early classification tasks as they want to see the full set of action dynamics extracted from a full video.

II. PROPOSED WORK

This is the block diagram of the system. It consists of high resolution camera, PC, serial to USB converter, Microcontroller PIC16F877A, power supply, LCD and buzzer.

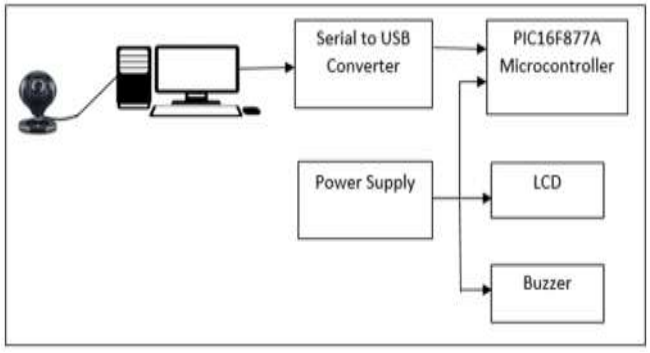


Fig. 1. Block Diagram

The image is captured by camera and transmitting to PC for processing. On PC there is MATLAB software. In this software, Image processing task is done.USB to serial converter is used to connect signals from PC to Microcontroller. The MATLAB generate control signals for microcontroller and these signals are sending via USB to serial converter. The power supply is given to the PIC microcontroller, LCD and buzzer.

Camera has following resolutions,

- Video Format: RGB 24 bit
- Video Resolution: 640x480, 1600x760, 1280x960, 1280x1024, 1600x1200, 2304x1728.
- Frame Rate: 30 Frames per second.

PIC 16F877A has following specifications,

- High-Performance RISC CPU, Only 35 single-word instructions to learn.
- Timer0: 8-bit clock/counter with 8-bit prescaler
- Timer1: 16-bit clock/counter with prescaler, can be augmented amid Sleep by means of outer gem/clock
- Timer2: 8-bit clock/counter with 8-bit period register, prescaler and postscaler

LCD display has following resolutions,

- Display Format: 16 Character x 2 Line
- Viewing Direction: 6 O'clock
- Input Data: 4-Bits or 8-Bits interface accessible

Display Font

- Display Font : 5 x 8 Dots
- Backlight LED.

USB to serial converter has following specifications,

- Single-Chip USB to UART Data Transfer
- USB Function Controller
- Virtual COM Port Device Drivers

A) Patch Aware Model

Given the representation of an interaction video frame. Our aim to figure out the interaction class and individual actions as well as analyze supporting regions for each interacting person[3]. We are given N training samples,

$$\{\mathbf{x}^{(i)}, \mathbf{p}^{(i)}, y^{(i)}\}_{i=1}^{N} \quad \ldots. (1)$$

Where $x \in R^D$ signify the video features, p = (p1, p2) is the individual actions of two interacting people, and $y \in Y$ is the interaction class.

Each hidden variable $h_i \in \{0,1\}$ is a binary variable signify that the *j*–th patch is correlated with background (h_j =0) or foreground (h_j =1).

Note that intra-class variability result in distinct patch formation in certain interaction classes.

An undirected group $G = (v , \varepsilon)$ is employed to encode the formations of these patches. A vertex $h_j \in v$ ($j = 1, ...M$) according to the *j*-th patch and an edge $(h_j, h_k) \in \varepsilon$ according to the dependency between the two patches.

We define the discriminative function as,

$$f(\mathbf{x}; \mathbf{w}) = \arg\max_{y,\mathbf{p}} \left[\max_{\mathbf{h}} F(\mathbf{x}, \mathbf{h}, \mathbf{p}, y; \mathbf{w}) \right] \quad \ldots.(2)$$

B) Support Vector Machine

Extracted features are passed to Support Vector machine in order classify event between two classes .Support Vector Machines are based on the concept of decision planes which define decision boundaries. A decision plane or decision boundary is one that isolates between arrangements of objects having different class. A schematic case is appeared in the representation beneath [4]. In this case, the items have a place either with class GREEN or RED. The separating line sets a limit on the right half of which all objects are GREEN and to one side of which all objects are RED. Any new object (white circle) falling to the right is labeled as GREEN and if it falling to the left of the separating line it is classified as RED.

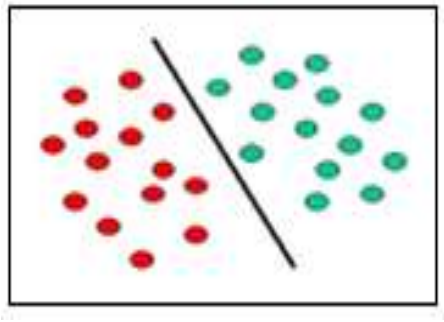


Fig. 2.Linear Classifier

The above is a simple example of a linear classifier, i.e., a classifier that separates a set of objects into two groups (GREEN and RED in this case) with a line. Most classification tasks, in any case, are not that straightforward, and regularly more complex structures are required so as to make an ideal division, i.e., accurately group new protests (test cases) on the premise of the cases that are accessible (train cases). This circumstance is portrayed in the outline underneath. Compared with the past schematic, plainly a full detachment of the GREEN and RED objects would require a bend (which is more complex than a line).Classification tasks based on drawing separating lines to differentiate between objects of different class memberships are known as hyperplane classifiers. Support Vector Machines are particularly suited to handle such tasks where linear classifier could not work.

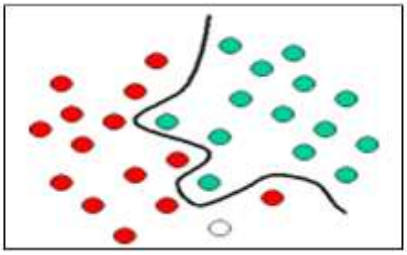


Fig. 3. Hyperplane Classifier

The basic idea behind Support Vector Machines. Here we see the original objects (left side of the schematic) are rearranged using a set of mathematical functions, which

known as kernels. The way toward rearranging the objects is known as mapping (change). Note that in this new setting, the mapped objects (right half of the schematic) is directly distinguishable and, subsequently, rather than developing the perplexing bend (left schematic), we should simply to locate an ideal line that can isolate the GREEN and the RED objects.

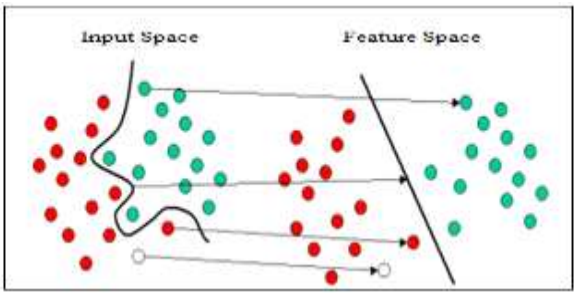
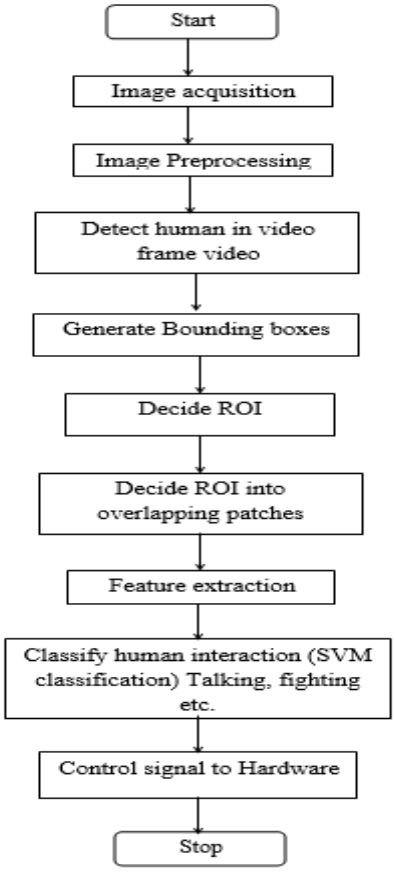


Fig. 4. SVM Classifier

III. OVERVIEW OF SYSTEM DEVELOPMENT

A) Software implementation

One of the major challenges in understanding human interactions is close physical contact, which often results in occlusion of body parts of interacting people. In close human interactions (e.g., push, hug, and hi-five), motion ambiguity significantly increases since commonly used features such as interest points and trajectories are difficult to be uniquely assigned to a particular person.



Therefore, recognizing individual actions becomes even more challenging in such interactions; and limits the performance of interaction classifiers. Previous work only considers interaction classes without close physical contact (e.g., handshaking, talking, and queuing) and uses a detector or tracker to extract each interacting person.

However, body part trackers and human detectors perform poorly when there are diverse categories of human motion that contain significant pose variations, limiting the performance of their interaction classifiers. Moreover, there are large variations in videos, including changes in subject appearance, scale, viewpoint, moving people and objects in the background [2][6], etc. These variations make the motion patterns of human interactions much noisier and thus a robust interaction recognition algorithm is required [5].

Fig. 5. Original Image

Video is nothing but contineous images. Fig shows the captured image from camera which is nothing but video frame. This RGB image is then converted in to the gray scale image and binary scale image.

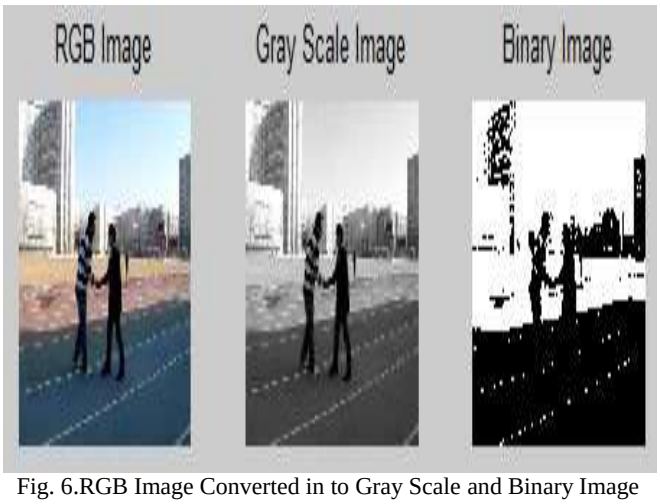


Fig. 6.RGB Image Converted in to Gray Scale and Binary Image

Figure shows the given RGB image converted in to the gray scale and binary scale image. In this we prefer the gray scale image for the further process. Memory required for the binary image is one bit, gray image is eight bit and RGB image is twenty four bit. So we use gray scale image for further process.

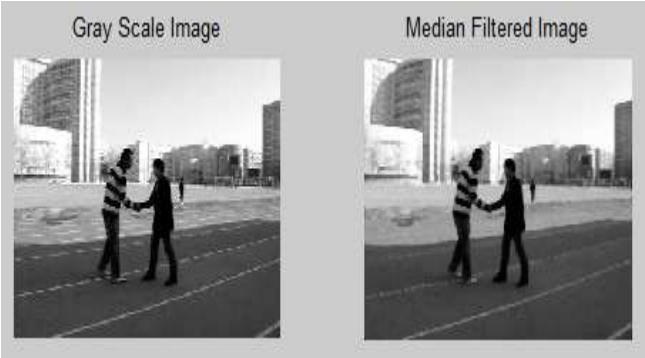


Fig. 7.Median Filter applied on Image to remove paper and salt noise

Most of the images contain paper and salt noise. To remove this paper and salt noise we use the median filter. The output after the removing noise in the image is shown. We use median filter because it has high PSNR. Edges are saved when we use median filter other than convolution.

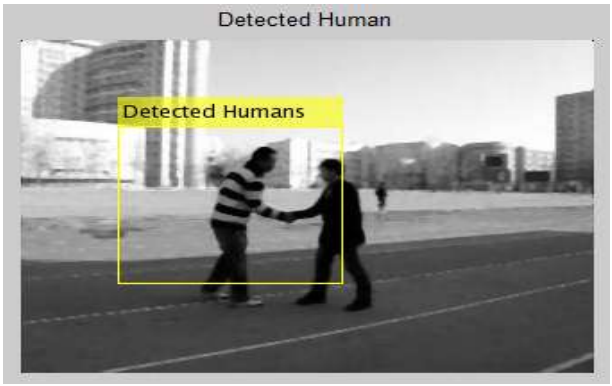


Fig. 8. Human Detected in Image

After improving the quality of the image we detect the human body in the image using the Matlab code. The output in the Matlab after detecting the human is shown in the fig which shows the patch with detected human.

B) Hardware implementation

In this we are interfacing the USB camera to the PC. Using USB to serial converter we connect PC to PIC microcontroller to send the signals to the hardware. The LCD display and buzzer are interfaced to the PIC microcontroller to show the outputs of the hardware.

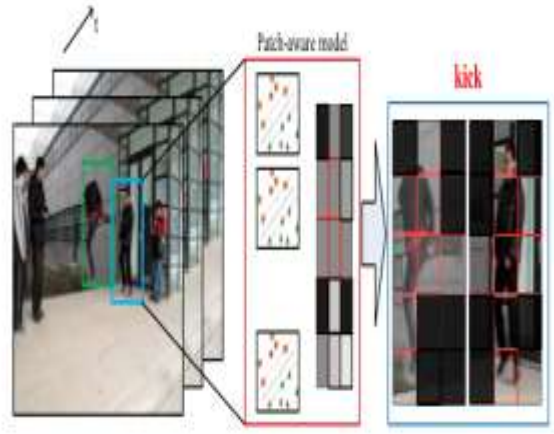


Fig. 9.Interface Results of Patch-Aware Model

Our model identify the human interactions and offers the ability to selectively learn the supporting regions can be used to separate the people from the background.

We separate out interacting people from video frame and obtain the volume for such person. Then the set of non-overlapping 3D patches are sampled from a volume. We use the advanced patch-aware model to conclude the interaction class. The labels of the latent patch variables can also be understood from the model, which shows whether the patches are associated with the background or foreground.

IV. RESULT AND CONCLUSION

We have intended a novel belonging to model for collectively recognizing interactions and individual actions from video frames. Our approach treats a video frame as a set of spatiotemporal patches. We utilize appearance as well as structural information of patches for finding the discriminative patches for classification. To show appearance feature of patches that are combined with the background, we built virtual video words for the patches and separate them from patches combined with people. Our model build upon the support vector machine in which patches are treated as latent variables which helps us with large variation of individual motion and human pose. Research show that our design achieves assuring results in interaction recognition.

REFERENCES

1] T. Lan, Y. Wang, W. Yang, S. N. Robinovitch, and G. Mori, “Discriminative latent models for recognizing contextual group activities,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 8, pp. 1549–1562, Aug. 2012.

2] A. Prest, C. Schmid, and V. Ferrari, “Weakly supervised learning of interactions between humans and objects,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 34, no. 3, pp. 601–614, Mar. 2012.

3] Y. Kong and Y. Fu, “Modeling supporting regions for close human interaction recognition,” in Proc. ECCV Workshop, 2014, pp. 29–44.

4] M. S. Ryoo and J. K. Aggarwal, “Recognition of composite human activities through context-free grammar based representation,” in Proc. CVPR, vol. 2. 2006, pp. 1709–1718.

5] A. Gupta, A. Kembhavi, and L. S. Davis, “Observing human-object interactions: Using spatial and functional compatibility for recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 10, pp. 1775–1789, Oct. 2009

6] Y. Kong, Y. Jia, and Y. Fu, “Interactive phrases: Semantic descriptions for human interaction recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 36, no. 4, pp. 1775–1788, Sep. 2014.

