STAR Research Journal

Available online at www.starresearchjournal.com (Star International Journal)

COMPUTER SCIENCE

Star. Computer. Science. 02 (2013)

HUMAN ACTION RECOGNITION USING FUZZY LOGIC ALGORITHM

S.Dhanalakshmi[1] and Dr.T.Ravichandran[2]

Associate Professor, Department of Computer Science & Engineering, SNS College of Technology,Coimbatore-641 035,India [1]

Principal, Hindusthan Institute of Technology, Coimbatore-641 042, India[2]

ARTICLE INFO

Article history:

Received: 21 Nov 2013;

Received in revised form: 26 Nov 2013;

Accepted: 30 Nov 2013;

Keywords

Computer Vision, Image Retrieval, Action Recognition

ABSTRACT

Action recognition in uncontrolled videos is a challenging task, where it is relatively hard to find the large amount of required training videos to model all the variations of the domain. This paper addresses this challenge and proposes a generic method for action recognition. Image is given as input where Action, position and motions are detected using fuzzy algorithm, in order to detect more accurately we use Histogram Equalization algorithm and stored in the Bitmap vector. When query is given it refer the Bitmap vector and displays the output for the query.

.

Introduction

Image processing is any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. Most image-processing techniques involve treating the image as a two-dimensional signal and applying standard signal-processing techniques to it. Image processing usually refers to digital image processing, but optical and analog image processing also are possible. This article is about general techniques that apply to all of them. The acquisition of images (producing the input image in the first place) is referred to as imaging. Modern digital technology has made it possible to manipulate multi-dimensional signals with systems that range from simple digital circuits to advanced parallel computers.

An image may be considered to contain sub-images sometimes referred to as regions-of-interest, ROIs, or simply regions. This concept reflects the fact that images frequently contain collections of objects each of which can be the basis for a region. In a sophisticated image processing system it should be possible to apply specific image processing operations to selected regions. Thus one part of an image (region) might be processed to suppress motion blur while another part might be processed to improve color rendition. The most requirements for image processing of images is that the images be available in digitized form, that is, arrays of finite length binary words. For digitization, the given Image is sampled on a discrete grid and each sample or pixel is quantized using a finite number of bits. The digitized image is processed by a computer. To display a digital image, it is first converted into analog signal, which is scanned onto a display.

Closely related to image processing are computer graphics and computer vision. In computer graphics, images are manually made from physical models of objects, environments, and lighting, instead of being acquired (via imaging devices such as cameras) from natural scenes, as in most animated movies. Computer vision, on the other hand, is often considered high-level image processing out of which a machine/computer/software intends to decipher the physical contents of an image or a sequence of images (e.g., videos or 3D full-body magnetic resonance scans).

In modern sciences and technologies, images also gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data). Examples include microarray data in genetic research, or real-time multi-asset portfolio trading in finance. Before going to processing an image, it is converted into a digital form. Digitization includes sampling of image and quantization of sampled values. After converting the image into bit information, processing is performed. This processing technique may be, Image enhancement, Image reconstruction, and Image compression

Image enhancement

It refers to accentuation, or sharpening, of image features such as boundaries, or contrast to make a graphic display more useful for display & analysis. This process does not increase the inherent information content in data. It includes gray level & contrast manipulation, noise reduction, edge crispening and sharpening, filtering, interpolation and magnification, pseudo coloring, and so on.

Image restoration

It is concerned with filtering the observed image to minimize the effect of degradations. Effectiveness of image restoration depends on the extent and accuracy of the knowledge of degradation process as well as on filter design. Image restoration differs from image enhancement in that the latter is concerned with more extraction or accentuation of image features.

Image compression

It is concerned with minimizing the no of bits required to represent an image. Application of compression are in broadcast TV, remote sensing via satellite, military communication via aircraft, radar, teleconferencing, facsimile transmission, for educational & business documents , medical images that arise in computer tomography, magnetic resonance imaging and digital radiology, motion , pictures ,satellite images, weather maps, geological surveys and so on.

RELATED WORK

Image Segmentation

Image Segmentation is the process of partitioning a digital image into multiple segments (sets of pixels, also known as super pixels). The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. Fig.1. Shows the original image with segmented image. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics.
The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image (see edge detection). Each of the pixels in a region are similar with respect to some characteristic or computed property, such as color, intensity, or texture. Adjacent regions are significantly different with respect to the same characteristic(s). When applied to a stack of images, typical in medical imaging, the resulting contours after image segmentation can be used to create 3D reconstructions with the help of interpolation algorithms like marching cubes.

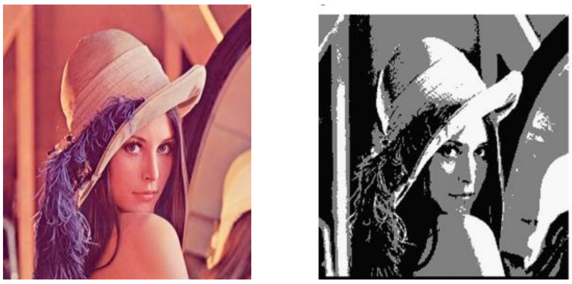

Fig.1.Original image with Segmented Image

Histogram equalization

Histogram equalization is a method in image processing of contrast adjustment using the image's histogram. This method usually increases the global contrast of many images, especially when the usable data of the image is represented by close contrast values. Through this adjustment, the intensities can be better distributed on the histogram. This allows for areas of lower local contrast to gain a higher contrast. Histogram is a graphical representation of data distribution. Histogram is used to enhance the image contrast or to determine the threshold values. Fig.2. Shows the Gray Scale Image with HOG graph.

Histogram equalization accomplishes this by effectively spreading out the most frequent intensity values. The method is useful in images with backgrounds and foregrounds that are both bright or both dark. In particular, the method can lead to better views of bone structure in x-ray images, and to better detail in photographs that are over or under-exposed. A key advantage of the method is that it is a fairly straightforward technique and an invertible operator.
The calculation is not computationally intensive. A disadvantage of the method is that it is indiscriminate. It may increase the contrast of background noise, while decreasing the usable signal. In scientific imaging where spatial correlation is more important than intensity of signal, the small signal to noise ratio usually hampers visual detection. Histogram equalization often produces unrealistic effects in photographs; however it is very useful for scientific images like thermal, satellite or x-ray images, often the same class of images that user would apply false-color to. Also histogram equalization can produce undesirable effects when applied to images with low color depth.

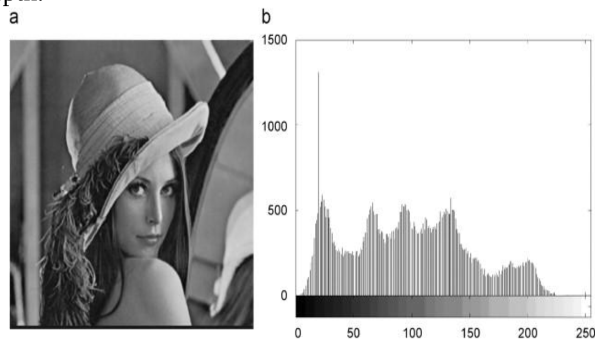

Fig.2.Gray Scale Image with HOG graph.

Fuzzy logic

Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based. Natural language (like most other activities in life and indeed the universe) is not easily translated into the absolute terms of 0 and 1. (Whether everything is ultimately describable in binary terms is a philosophical question worth pursuing, but in practice much data we might want to feed a computer is in some state in between and so, frequently, are the results of computing. Fuzzy logic includes 0 and 1 as extreme cases of truth (or "the state of matters" or "fact") but also includes the various states of truth in between so that, for example, the result of a comparison between two things could be not "tall" or "short" but ".38 of tallness."
Fuzzy logic seems closer to the way our brains work. We aggregate data and form a number of partial truths which we aggregate further into higher truths which in turn, when certain thresholds are exceeded, cause certain further results such as motor reaction. A similar kind of process is used in artificial computer neural network and expert systems.It may help to see fuzzy logic as the way reasoning really works and binary or Boolean logic is simply a special case of it.
Fuzzy logic is a form of many-valued logic or probabilistic logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values) fuzzy logic variables may have a truth value that ranges in degree between 0 and 1. Fuzzy logic has been extended to handle the concept of partial truth, where the truth value may range between completely true and completely falls. Furthermore, when linguistic variables are used, these degrees may be managed by specific functions.

Smoothing

It is a fundamental operation in image processing so is the ability to take one or more spatial derivatives of the image. The fundamental problem is that, according to the mathematical definition of a derivative, this cannot be done.

A digitized image is not a continuous function a(x,y) of the spatial variables but rather a discrete function a[m,n] of the integer spatial coordinates. As a result the algorithms we will present can only be seen as approximations to the true spatial derivatives of the original spatially-continuous image. Further, as we can see from the Fourier property takes a derivative multiplies the signal spectrum by either *u* or *v*.Fig.4. Original Image with Smoothing Image. This means that high frequency noise will be emphasized in the resulting image. The general solution to this problem is to combine the derivative operation with one that suppresses high frequency noise, in short, smoothing in combination with the desired derivative operation. As an image is a function of two (or more) variables it is necessary to define the direction in which the derivative is taken. For the two-dimensional case we have the horizontal direction, the vertical direction, or an arbitrary direction which can be considered as a combination of the two.

Fig.4.Original Image with Smoothing Image

Tracking

Each frame can depict multiple people. Therefore, some form of data association is needed to keep track of each person's appearances from frame to frame. We adopt a simple bounding box based tracking procedure for obtaining tracks of each person. We do this by initializing the tracker to the detections in the first frame. In subsequent frames, each new detection is added to the previous person track that has the closest spatial position. If detection cannot be associated with one of the existing tracks, a new track is initialized. This simple scheme has several, well-known weaknesses. For example, it may easily fail if the camera is moving extremely fast. However, it turns out that this method provides sufficient accuracy (with a few extra tracks) for our purposes.

Visual Tracking

We presented the notion of an image-domain landmark as a local maximum of edge density. A landmark represents the basic feature which we employ for localization, a task which will be accomplished using a characterization of the landmark's appearance as a function of the camera's position in configuration space. In order to achieve this characterization, however, the landmark must first be tracked. A qualitative analysis of the results shown in however, indicated that landmark candidates do not necessarily correspond precisely from one image to the next. This chapter will explore the problem of tracking in general and present our particular approach to the problem, given the input generated by the landmark detector. In computational vision, visual tracking is the act of consistently locating a desired feature in each image of an input sequence. The problem is typically complicated by sensor noise, motion in the scene, motion on the part of the observer and real-time constraints. The problem can be further complicated when more than one identical feature must be tracked, in which case it is up to the observer to decide the optimal set of correspondences which are consistent with a priori assumptions about, and recent observations of, the behavioral characteristics of the features.

Our technique for landmark tracking operates as follows. Given an initial set of prototypes, that is, observations of a set of unique landmarks, a tracked landmark, is constructed for each prototype. A tracked landmark is constructed by identifying matches to its prototype amongst the set of all observed landmark candidates. In practice, since landmark candidates can demonstrate local variation in position as the camera moves, a local search in the image neighborhood of a candidate may be required. We will refer to the task of matching a single candidate landmark to a prototype as landmark recognition, and the task of building tracked landmarks as landmark tracking. Provides an overview of the training process presented thus far; candidate landmarks are detected as local maxima of edge density and then tracked into sets of tracked landmarks. Outlined the process of candidate extraction, while the following sections will present the tasks of landmark recognition and tracking over multiple images.

Multi-scale segmentation

Image segmentations are computed at multiple scales in scale space and sometimes propagated from course to fine scales; see scale-space segmentation. Segmentation criteria can be arbitrarily complex and may take into account global as well as local criteria. A common requirement is that each region must be connected in some sense. One-dimensional hierarchical signal segmentation: Witkin's seminal work in scale space included the notion that a one-dimensional signal could be unambiguously segmented into regions, with one scale parameter controlling the scale of segmentation.

PROPOSED WORK

3.1 System Architecture

The system architecture consists of an image as the input image, camera image and the binary image and the image undergoes the process of segmentation. Then the segmentation process should perform the either binary image or face recognition. Then its perform the segmentation and determination of the body of regions. Fig.5.Shows the system architecture of segmentation process.

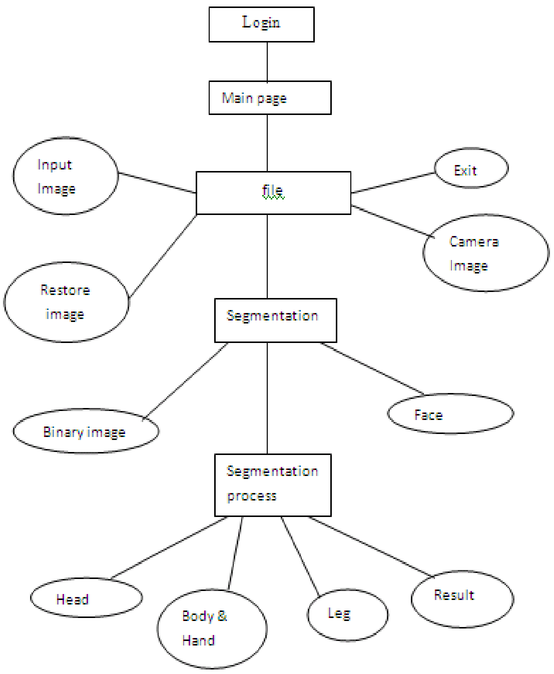

Fig.5. System Architecture

EXPERIMENTAL RESULTS

4.1Action Detection from Input Image

For Action detection, first convert binary image from RGB image. For converting binary image, calculate the average value of RGB for each pixel and if the average value is below than 110, replace it by black pixel and otherwise replace it by white pixel. By this method, binary image from RGB image is extracted. Fig.6. Shows the Input Image of an action detection.

Find the upper portion from the binary image. Start scanning from the middle of the image, then finds continuous white pixels after a continuous black pixel. Find the maximum width of the white pixel by searching vertical both left and right site. Cut the face from the starting position of the forehead and its high will be 1.5 multiply of its width.

Fig.6.Input Image

4.2 Segmentation & Determination of the body Region

This module uses the top and bottom region, first represent the image in the L * a * b space from its conventional red–green–blue (RGB) space. The L * a * b system has the additional benefit of representing a perceptually uniform color space. It defines a uniform matrix space representation of color so that a perceptual color difference is represented by the Euclidean distance. The color information, however, is not adequate to identify the hand region. The position information of pixels together with their color would be a good feature to segment the hand region from the body. The Fuzzy C-means clustering algorithm that employ to detect the top region is supplied with both color and pixel-position information of the image. This module is used in image segmentation in general and hand region segmentation in particular is a novel area of research. A description of the FCM clustering algorithm can be found in books on fuzzy pattern recognition. Fig.7. Shows the determining the region of an image segmentation. Determination of MO in a black and white image is easier because of the presence of the hand region

A plot of the average intensity profile against the MO reveals that the curve has several minima, out of which the first and third correspond to the inner region of the top and the inner region of the bottom region of body, respectively. The difference between the preceding two measurements along the Y-axis gives a measure of the MO. Fig.8. Shows the determining the binary region of an image segmentation.

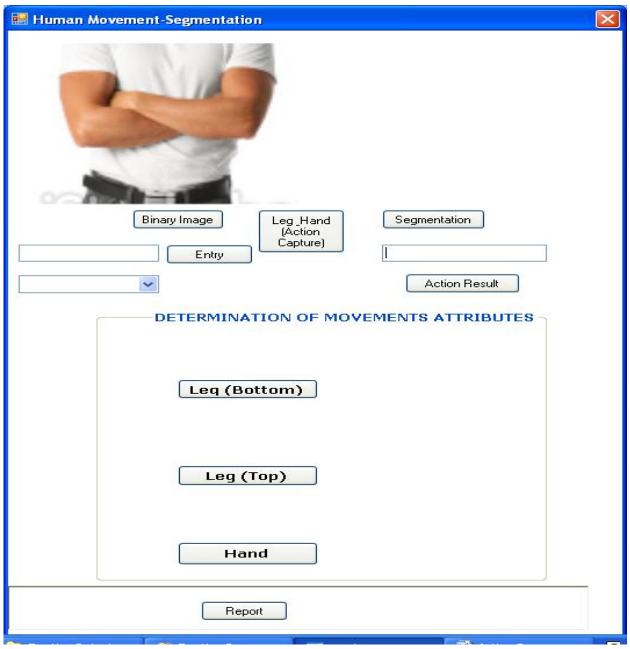

Fig.7.Determining the region

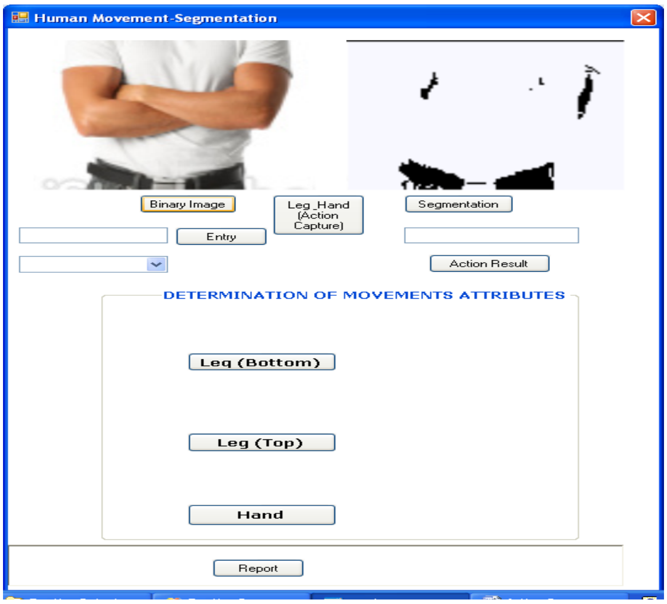

Fig.8.Determining the binary image

4.3 Segmentation & Determination of the bottom Region

The bottom region in a monochrome image has a sharp contrast to the rest of the body. Consequently, the thresholding method can be employed to segment the leg region from the image. Images grabbed at poor illumination conditions have a very low average intensity value. Segmentation of the leg region in these cases is difficult because of the presence of motion in the neighborhood of the leg region. To overcome this problem, consider images grabbed under good illuminating conditions.

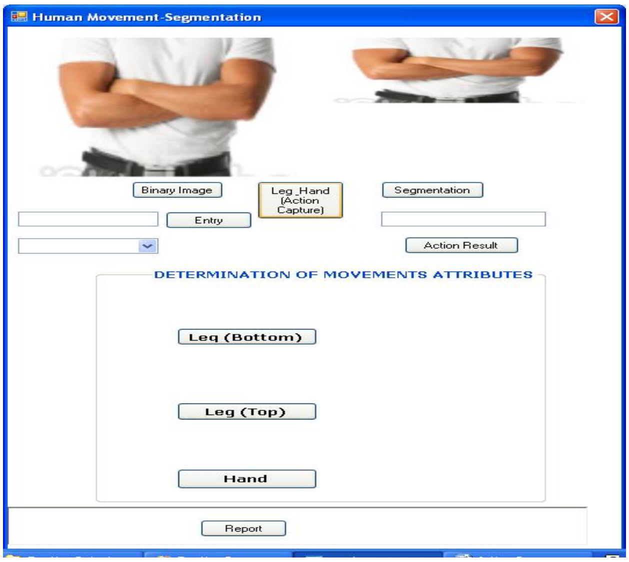

Fig.9.Image Segmentation

Fig.9. shows the image segmentation of an different region. After segmentation of the image, localize the left and right leg on the image. In these Modules, template matching scheme is used to localize the leg. The leg template we used is similar to the template. The template matching scheme, taken from our previous works, attempts to minimize the Euclidean distance between a fuzzy descriptor of the template and the fuzzy descriptor of the part of the image where the template is located.Fig.10. Shows the movement determination of an input image. Even when the template is not a part of the image, the nearest matched location of the template in the image can be traced.

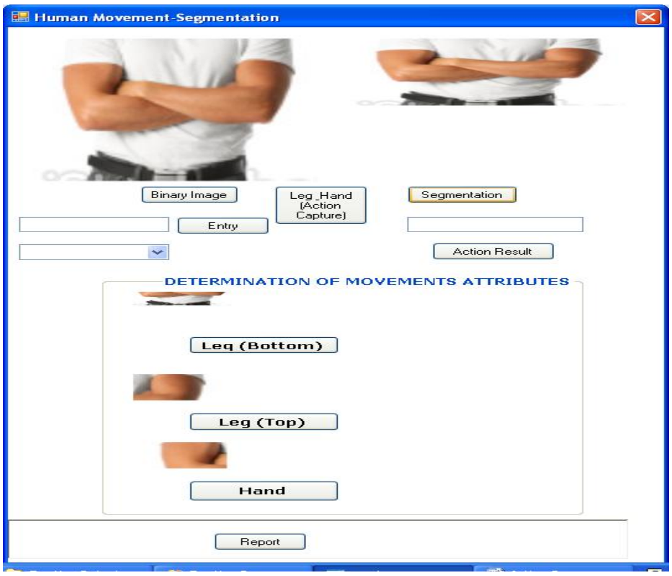

Fig.10.Movement determination

4.4 Localization & Determination of EBC Region

In the image, hand region is easily segmented by setting a very low threshold in the histogram-based thresholding algorithm. The hand regions are also segmented by thresholding. A search for a dark narrow template can easily localize the hand. Note that the localization of the hand is essential for determining its length. Constriction in the hand region can be explained as a collection of white and dark patches called hilly and valley regions, respectively. The valley regions are usually darker than the hilly regions. Fig.11. Shows the movement determination of an region.

Fig.11.Determining the position

4.5 Action-Recognition

For action detection of an image, we have to find the Bezier curve of the, left leg and right leg. Then we convert each width of the Bezier curve to 100 and height according to its width. If the person's action information is available in the database, then the program will match which emotion's height is nearest the current height and the program will give the nearest action as output. Fig.12. shows the action recoginition of an input image.

Fig.12.Action Recognition

CONCLUSION

The proposed OPP descriptor takes one step further and models the temporal relationships between poses. In this way, actions that share similar intermediate poses can be more accurately discriminated. There is room for improvement. Action image retrieval brings a set of challenges. First, the data retrieved is quite noisy and consists of multiple modes due to the variance in poses and in people. This makes the problem more challenging and requires special attention. While NMF is successful in finding modes of the data to a degree, it may face some problems such as slow convergence for large datasets or having no guarantee for the global optimum. For this case, mode finding can be improved by using additional techniques like nonnegative patch alignment. Second, from the retrieval point of view, the regular cues (like color, static templates) used in content-based retrieval of objects are likely to fail in this domain. We therefore use HOG features operating on Pb responses for describing action images. Additional pose cues will likely improve the performance. Note that, any performance increase in the action image retrieval step is likely to improve the further steps of the recognition. On the other hand, the retrieved data is quite diverse, and using this data effectively can be very beneficial. We have seen that the action images are also composite in nature (running and waving, for example), like the actions in video. It would be interesting to discover the most influential ordering pairs for specific action. The pose pairs can be either that occurs together or should not occur together. Also detailed analysis relationship can further help recognition.

REFERENCES

[1] Nazli Ikizler-Cinbis and Stan Sclaroff(AUGUST 2012),, "Web-Based Classifiers for Human Action Recognition" IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 14, NO. 4.

[2] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri,(2005) "Actions as space-time shapes," in Proc. ICCV,vol. 2, pp. 1395–1402

[3] N. Ikizler-Cinbis, R. G. Cinbis, and S. Sclaroff,(2009) "Learning actions from the web," in Proc. Int. Conf. Comput. Vis, pp. 995–1002.

[4] Schuldt, I. Laptev, and B. Caputo,(2004) "Recognizing human actions: A local svm approach," in Proc. ICPR,pp. 32–36.

[5] D. Weinland and E. Boyer,(2008) "Action recognition using exemplar-based embedding," presented at the CVPR, Anchorage, AK.

1. Angela D'Angelo, Li Zhaoping, and Mauro Barni(APRIL 2010) "A Full-Reference Quality Metric for Geometrically Distorted Images" IEEE transactions on image processing, VOL. 19, NO. 4,

2. C. Schuldt , I. Laptev and B. Caputo(2004) "Recognizing human actions: A local svm approach", Proc. ICPR, pp.32 -36

3. C. Thurau and V. Hlavac (2008) "Pose primitive based human action recognition in videos or still images", CVPR.

4. D. Tran and A. Sorokin (2008) "Human activity recognition with metric learning", Proc. ECCV, pp.548 -561

5. D. Weinland and E. Boyer (2008) "Action recognition using exemplar-based embedding", CVPR.

6. H. Jhuang , T. Serre , L. Wolf and T. Poggio (2007) "A biologically inspired system for action recognition", ICCV

7. J. C. Niebles , H. Wang and L. Fei-Fei (2006)"Unsupervised learning of human action categories using spatial-temporal words", Proc. BMVC, pp.1249 -1258

8. K. Schindler and L. van Gool (2008) "Action snippets: How many frames does human action recognition require?", CVPR.

9. M. Blank , L. Gorelick , E. Shechtman , M. Irani and R. Basri(2005) "Actions as space-time shapes", Proc. ICCV, vol. 2, pp.1395 -1402

10. N. Ikizler-Cinbis , R. G. Cinbis and S. Sclaroff (2009) "Learning actions from the web", Proc. Int. Conf. Comput. Vis., pp.995 -1002.

11. Nazli Ikizler-Cinbis and Stan Sclaroff, (AUGUST 2012) "Web-Based Classifiers for Human Action Recognition" IEEE transactions on multimedia, VOL. 14, NO. 4.

12. T. K. Kim , S. F. Wong and R. Cipolla (2007)"Tensor canonical correlation analysis for action classification", CVPR.