DOI: http://dx.doi.org/10.26483/ijarcs.v9i1.5287

ISSN No. 0976-5697

Volume 9, No. 1, January-February 2018

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

HUMAN GESTURE ANALYSIS BASED ON VIDEO

Snowber Mushtaq
Department of Computer science
National Institute of technology Srinagar
Srinagar, Jammu and Kashmir, India

Tausifa Jan Saleem
Department of Computer science
National Institute of technology Srinagar
Srinagar, Jammu and Kashmir, India

Abstract: Human gesture recognition is defined as a necessity to determine what human body actions occur in videos. Gestures can originate from body motion like walking, bending, jumping, and hand waving. When a video is playing the human action detection is a difficult point of detection. This problem is peculiarly hard due to extensive variations in motion appearance of actions, camera angles with respect to the human body, motion in the background, noise and large amount of video data. Main aim is to detect various gestures in a multimedia clip by pre-processing the video and then applying algorithm for identifying various actions. The important role is to determine behavior of humans based on based on their actions. The moving objects are determined from a video stream. The database used includes CASIA database and WEIZ MANS ACTION database to test the proposed system. Those can be applied to Surveillance Systems.

Keywords: silhouttee, behavior analysis, action recognition ,skiltonization..

I. Introduction

A gesture is a form of action based communication in which visible human body actions are seen and they can be a way of communication and can generate a particular messages. The various movements included in gesture are the movement of the hands, legs, arms, face, or other parts of the body. The actions generated include arm stretching, walking, leg stretching, jumping, bending, skipping etc. They differ from real non vocal expression that does not communicate some particular messages, but express some information such as completely expressive and exhibits of complete attention. Individuals are allowed to communicate a variety of actions by body movements.

We know the automatic analysis of surveillance video is nowadays major field of investigation in computer industry and research. This field holds the techniques that can interpret the behaviour of the human from scene during various situations. The major step is to detect situations of our interest and particularly when some unexpected thing occurs. The aim is to automatically interpret the scene and detect the abnormal behaviour. We know that from last two decades the surveillance cameras has increased and most of the places like hospitals, colleges, airports etc. cameras are installed. The main reason of installation of cameras is fear of crime. Traditionally, the video streams from cameras are monitored by humans who are trained to monitor the abnormal spot and the alternatives is to store these in the database.

The interest area is automatic identification of real objects or recorded objects. The human body motion investigation can be applied in a variety of field domains, such as video regeneration, video surveillance, human-computer interaction systems, and medical analysis. Motion tracing of human is a significant and challenging computer problem. It covers the detection, tracking, and perhaps understanding of human actions in the videos or image sequence. In some cases, people with suspicious behavior are analyzed and they are identified and their unusual behavior are directly detected from videos.

There are many applications in computer vision and image processing which require human identification in a video stream and understanding of their actions through automated processing. The authors are currently developing a project that focus on providing such functionality, additionally with gesture recognition and interpretation. The system that is being developed called moveIt (movements Interpreted) is an automated system to identify, track, recognize, interpret, and analyse whole-body type gestures to determine the behaviour of humans from a video stream. For instance, imagine an airport where a security officer a in control room visually checking all the video feeds from CCTV cameras fixed in several parts of the airport to make sure that no suspicious or criminal activity is occurring. If he detects anything unusual (or some suspicious behaviour), he immediately notifies the other officers closer to that area to take appropriate action. Performing such a task visually is tiresome, boring and carries significant chance that the officer may miss to detect an image of interest. Next consider the possibility of automating this task where a machine does the complex work of monitoring, analysing and determining activities of interest and doing it intelligently. The moveIt project focus on building such a system.

The applications of human action recognition include elderly-caring, child care, various disease can be found like ADHD (Attention-deficit hyperactivity disorder). ADHD is a cause of sleep disorder found in children below age of eight. The child with ADHD disease have frequent body movement during sleep as compared to normal children. Therefore, investigation is to be needed to find body movement during sleep of children with and without ADHD disease using video imaging. The video from cameras can weather a human is trying to pass when a car is moving so that accidents can be prevented. Human gesture and action recognition systems focuses to recognize the actions of one or more humans from a continuous observations of the videos. It is one of the main aspect of several applications such as visual surveillance, video intent and interaction of computer interaction with human among. Perception of human activities is the last step

© 2015-19, IJARCS All Rights Reserved

before that various steps included can be video capture, frame extraction, silhouette extraction, tracking, identification, and classification of various actions. The action and activity recognition an also include motion analysis understanding various scene and activities of humans from those scenes, understanding human actions, classifying those human actions or human body motions from captured video.

2. LITERATURE REVIEW

In Skeltonization of real time gesture recognition system [1], the idea is automatic identification of real time objects or recorded objects from a video and then to analyze those objects. First we find the moving objects from video and then to find whether the moving object is human or non-human. The human object actions need to be identified through automated processing. The tracking of human object in different frames can be done using contour and blobs. A group of connected pixels in an image that have shared certain common characteristics is called Blob E.g. their color and intensity values are same. If in the image we have the dark connected region that is blobs, and the aim is detection of blobs, identify and mark these regions. Now, for the detection of blobs we have to do: (a) converting source image to binary image based on certain threshold. We need to find certain minimum threshold and on the basis of this threshold the values below that are converted into black or 0 and values above that are converted into white or 1 and this process is known as thresholding. (b) In the above binary image the connected white pixels are grouped together and are called binary blobs and this process is known as grouping. (c) Then the center of the binary blobs are computed, the area of blob closer this by a certain minimum distance are merged together and the process is known as merging. (d) Last step is to calculate the center and radius of the new blob by merging process and this is known as center and radius calculation. The filtering of blobs can be done by color, size or shape. (a) Filtering blobs based on color, in this blob_ color =0 to select darker blobs and blob_color =255 for lighter blobs. (b) Filtering on the basis of size, in this various parameters include (a) circularity: Here we select the circular blob e.g. the circularity of regular hexagon is more than rectangle .The circularity of a circle is maximum and is equal to 1 and it is .78 in case of rectangle. (b) Convexity: Convexity can be defined as the (Area of the Blob / Area of its convex hull). Convex Hull of a shape is the tightest convex shape that completely encloses the shape.

The movements in various frames need to be modelled and generalized in order to recognize the gestures. The model of the human is in the form of a Skelton. In actual sense we need faster Skeltonization and at the same time results should be accurate. So, for that it has introduced Star Skeltonization and is considered one of the basic methods for Skeltonization. The basis of this algorithm is to finding the five extreme points of the Skelton. Because of the five extreme points and connecting to the center it resembles a star and so it is known as star Skeltonization. The borders of the five extreme points are of the human model are selected. Then the center point of the overall model is calculated by finding the average point of all the borders .Next graph is plotted between boundary and centroid. The plotted graph is smoothened using low pass filter or Fourier transform. From this smoothened graph the extreme points can be find out. Then these five extreme points are connected to the centroid and form the star Skelton.

The commonly recognized used method in three dimensional modelling is Curve-skeleton for computer vision. The thinned one dimensional are curve-skeletons and represent the 3D objects, and are useful for various applications such as reduced-model formulation, visualization improvement, animation. The method used is voxel topology, computational geometry. They give more control over matching algorithm for users .They are more common in areas like medical image diagnosis, that is the diagnosis of disease from images, computer graphics, computer aided design(CAD), remote sensing. The models based on thinning and curve-Skelton are similar to Skeltonization.

The morphological operation that is used to eliminate the pixels in foreground in a binary image is known as thinning.Thinning is the similar operation like erosion and dilation [2].It maintains the topology and object shape while skeletons are formed. The skeletons formed have width of one pixel. This is one of the good method in noise and shadow free environment. The proper Skelton cannot be formed if the parts of object are disconnected. The template matching can be performed in small regions of image, it reduces the computational time.it is biased towards areas where motion is produced.

In pixel density based star Skeltonization algorithm, the algorithm operates in three distinct steps for finding five extreme points i.e., finding pixel points locating head, two legs and two hands. The boundary box technique is prone to errors, so there is need for identifying different parts separately. First part of algorithm is identification of the head, this method assumes that the upper most part of the bounding box is the head. This assumption can be wrong only when a person bends and at that time the top most pert is not head as this is considered the special case for the identification of head position. The upper most section is divided into equal intervals and the pixel densities are calculated. To find the exact location of the head a graph is drawn between pixel densities and interval points and the shape of the graph is bell shaped and is the desired shape for the location of head. The noise will not affect the shape of the graph but there will be little deviations, hence is considered the accurate and robust one. And it can be applied to the real videos because of its accuracy and fastness.

Second one is the identficati on of hands, this also uses the same idea as for the identification of head, but data analysis method is used for locating the actual coordinate's .the images with background subtracted the hand positions cannot be calculated in case of two situations. (a)When the hands are lapped over the body of the human. (b) When the hands are behind the body. In the above two cases the position of hand are very difficult to compute from the calculated binary image. So the algorithm given rejects the given two cases. From the plotted graph it is observed that the pixel density increases with hand and the movement of this pixel density can be observed so that the movement of hands can be find out and direction of movement can also be find out. But abrupt motions of the hand cannot be predicted using the method. In this way we do not receive the graph so the computation is in efficient and under real-time processing of videos. But it has reasonable results under different cases to test the system.

Third is identification of legs, the actual leg coordinates is very difficult to find as the frames extracted get mixed with

© 2015-19, IJARCS All Rights Reserved

noise and the shadow of the person, which is difficult to leave out. So, the pixel based method is needed in the restricted area. The area calculated will be less than h/2 and h is the height of the person. But in case when a person bends, then identification of legs is a challenging task. Here we need to view frames back and forth for the actual positions of legs.

The disadvantage of the pixel density based method is that it cannot find certain postures like crawling which is a non-standard posture type and is thus less accurate for human images to find all postures .This method needs a lot of pixel based calculations which are simple ,robust and can be applied to real time video streams. The main idea is to find five extreme points and connecting it to the main body that is centroid. The distance from the extreme points to the centroid is calculated .The noise reduction should be applied to the distance graph by using low pass filter or smoothing filter.

A framework for whole-body gesture recognition from video feeds [3], in this the first step is to find the moving objects from the video. In order to find this the background subtraction is needed. The background subtraction method is mostly used for foreground extraction. And is found to be frequently used step in image processing [4]. The accuracy of this method depends on how it is able to detect the moving humans from a video. There are various algorithms available for the background removal but Adaptive Median background subtraction algorithm is chosen because it is high performance, memory efficient and very high quality output [5].The binary video is extracted from the original video. But the result binary video may contain noise and the shadows.

Shadow Detection and Removal, A shadow is an area where light from a light source is blocked by an object. All of the three-dimensional volume is occupies behind an object with light in front of it. A shadow is a two-dimensional silhouette, or reverse projection of the object blocking the light. The shadow is the main problem in background subtraction. So there is a need of an algorithm that can detect shadows and remove shadows and foreground of an image is extracted. The shadow has a similar chromaticity and has brightness lower .here shadow detection algorithm is based on the color. This color detection base algorithm is based both on chromaticity and color .The non-background and background pixel values are compared, and a threshold value is set. The values below this threshold are considered as the shadow and above it are considered as the object.

The Brightness distortion (BD) is defined as a value and brings expected background close to the observed chromaticity value. Color distortion (CD) can be defined as the orthogonal distance between the expected color and the observed chromaticity value. If the CD is less than 10 and BD is between .5 and 1 then it is shadow. But if CD is less than 10 and BD is between 1 and 1.25 then it is foreground. Thus this way we are able to classify shadow and non-shadow components of an object. Next we need to find out whether the foreground object extracted is the actual object of interest.

In object tracking, the moving objects are separated from the scene by applying background subtraction. The co-ordinates of the identified objects are find in all the incoming frames. Here first initialization occurs, next updating occurs lastly termination of tracking process occurs. Various tracking methodologies are available as per the application e.g. gesture recognition, face recognition .Here video is taken from a static camera and the object that is in motion is taken away by the background subtraction .Now before tracking ,the identified object is labelled and bounding box algorithm is performed. The bounding box in different frames is compared and associativity between frames is find out. The velocity of the objects is recorded in different frames and can find out position of object in current frame. In case where the object moves slowly the object velocity can be neglected.

The object tracked need to be modelled, so that actions can be recognized in consecutive frames. The model formed must be a general one so that all objects can be compared with that model. For this human Skelton is the best model, an identification of Skelton and comparing it with actions is not an easy task. A lot of processing is needed for high accuracy. The advantages of the above method is that it is able to process on real time videos and is computationally efficient. The main things needed for this is that, it should be taken from a static camera and it should be illuminated. The shaking of cameras will create significant amount of noise.

The various applications of the method include designing of safety systems for buildings. It can be used to identify the psychology of the children, the children that are quit and less interactive can be identified. The abnormal behavior of the children can be studied. This system can assist drivers while driving, it can identify persons while driving so that accidents can be avoided.

In tracking, analysis, and recognition of human gestures in video [6], the main aim is to detect gestures automatically and to recognize that from videos by applying video and image processing. The main aim is to locate, track hands in different images of a video .It estimates the hand shapes, the head gestures and various upper body motions in videos. It find the different gestures from video and summarizes those gestures. The algorithm implemented has been demonstrated for three different areas and those are hand signaling that are employed at airports, gesture based interaction for disabled persons and American Sign Language. The algorithms developed is general one and can be applied in various fields.

The head and facial gestures can be applied to the areas where people are paralyzed and are suffering from multiple strokes and this allows them a friendly approach to interact with their friends, family. AMERA MOUSE system [7] a computer user's movements are tracked with a video camera and then they are translated into the movements of the mouse pointer on the screen of a computer. The Features of the body like tip of the nose of a person can be tracked. The tracked feature that is tip of nose is initialized and normalized correlation coefficient is calculated to find the best search.

The upper body gestures can be used in signals of reference for various events like flight directions on the airport, the traffic police showing directions .Here machine learning framework is used to identify the various body gestures.

In video-based human movement analysis and its application to surveillance systems [8], posture classification system is presented and in this human movements are directly analyzed from video. Here every sequence is converted into a postures. For characterization of postures, triangulation method is used to divide into triangular meshes. From this two features are extracted that are Skelton features and centroid feature. The Skelton feature gives the representation of the object and centroid feature gives description of the object. The depth first search (DFS) is employed to extract the Skelton feature of a posture. The Skelton feature extraction is efficient and robust. The centroid is extracted from the Skelton feature. The two feature together gives human feature extraction more efficient and accurate for classification of postures. The posture are classified by string symbols.

3.PROPOSED METHODOLOGY

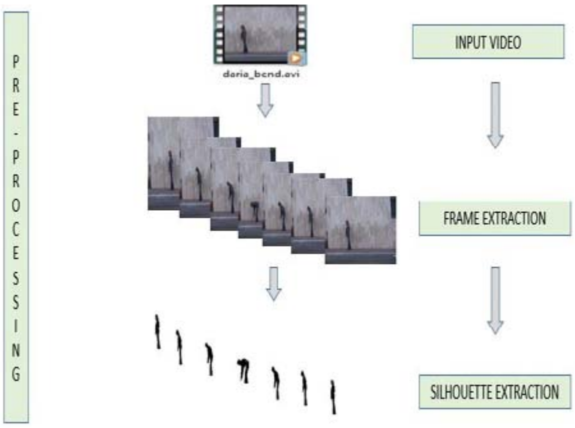
The aim is to implement the algorithm with can identify various gestures in a video. The steps include various processing steps on incoming video, the applying the algorithms to identify the particular gesture. The gestures identified are walking, bending, arm-stretching.

Main Steps Involved Are:

1) Pre-processing of videos.
2) Identification of gesture.

Step 1: Pre-Processing Of Videos

In this first the incoming video stream is read, the one by one frames are read and are copied to the output folder and is known as Frame extraction of video. Next step is to extract the silhouette from the individual frames. This is done by converting color image into grey images and then applying threshold based on certain value so that a person silhouette is extracted. By applying thresholding binary images are extracted. The binary image is the image containing logical values i.e. 0's and 1's.



Step 2: Applying Gesture Algorithms.

The gestures applying for different algorithms are

1. Walking Gesture:

The main idea is that the distance between the two foots change from maximum to 0 and then back to maximum when a person walks. There is one minimum between every two maxima. By this we can find the legs are stretching but we can't say a person is walking .In order to find whether the stretching is due to walking, we need to chance in position of one foot. By observing both two conditions we can say a person is walking.

Algorithm

```
• From 1 to no. of  frames
• Read each frame
• Calculate x and y co-ordinates of each frame
• Scan from bottom each pixel from left to right ,   find
    ycod_foot= ycod_of_bottom first white pixel
• Decrease ycod
    ycod = ycod - 5
• From left to right in the same ycod read each pixel  and find
    start_foot1=first white pixel
    End_foot1=first black pixel after start_foot1
    Start_foot2=first white pixel after end_foot1
    End _foot2=first black pixel after start_foot2
• Result=end _foot1 - start_foot2
```

The Result values are plotted with respect to the frame no's .The graph can be smoothened using smooth filter for better visibility of graphs. And the position of start_foot1 is plotted in different frames, it is observed that it is linear.

1. Bending Gesture:

In bending the height of a person changes, when a person starts bending the height first starts decreasing to the minimum and then after it increases .This can be analyzed by assuming the top most part is the head and finding the position head, the height in various frames are calculated and plotted.

```
• From 1 to no. of frames
• Read each frame
• calculate x and y co-ordinates of each frame
• scan from top pixel by pixel from left to right find
    First black pixel which indicates head
•  Save head positions of each frame.
```

3.ARM STRETCHING:

In arm stretching the main idea is that the left side co-ordinates change

```
• From 1 to no. of frames
• Read each frame
• calculate x and y co-ordinates of each frame
• scan from left to right  each pixel  and height > h/2  find
  black pixel in each frame which indicates position of arm.
```

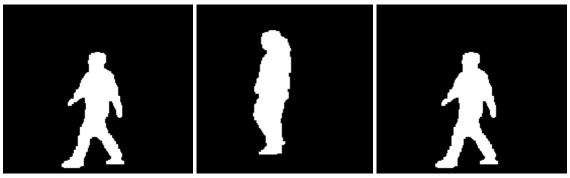
4. RESULTS

We conducted a set of experiments, using CASIA data set and the wiz man's action database.

A)The walking gesture was performed on CASIA database and it was observed that the maxima's i.e. the maximum no of pixels between the two foots were calculated in every frame and the maxima's arise when the legs are full stretched and minima's arise when the foots are closer. Between every two maxima's there is one minima.

Table 1: List of maxima's and minima's in different frame nos for different persons.

	Frame no's for maxima's and minima's						
Person 1	maxima's	6	19	32	45	58	71
	minima's	13	24	39	50	65	75
Person 2	maxima's	11	24	37	50	62	71
	minima's	4	18	29	44	56	75
Person 3	maxima's	10	22	34	46	59	67
	minima's	1	17	27	41	51	65



(a) (b) (c)

Figure 1(a) Maximum pixels between feet.1 (b) Minimum pixels between feet.1(c) Maximum pixels between feet.

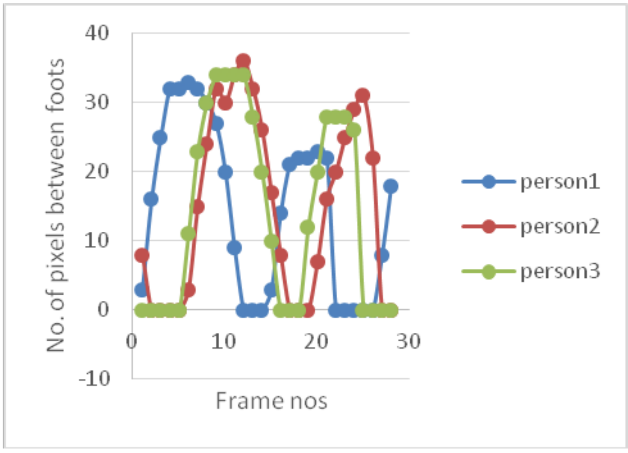


Figure 1: Variation of pixels between foot with frame no's for different persons.

The change in foot distance indicates that there is stretching of legs only but in order to find where the person is walking we need to find the position of the any of the foot with respect to their frame no's.

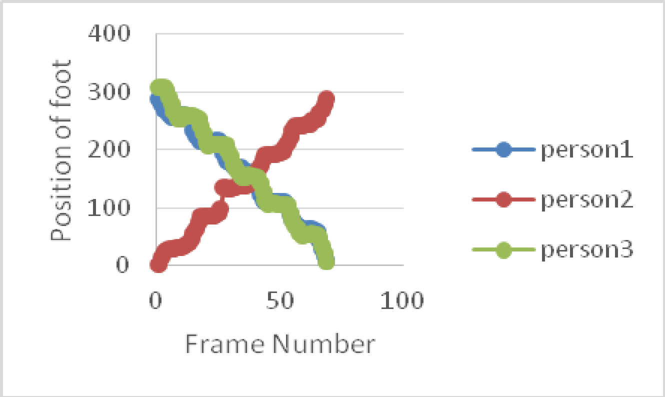


Figure 2:list of positions of foot in various frames for different persons.

Figure 3: Variation of foot position with frame no's for different persons.

Frame Nos	***Person1***	***Person2***	***Person3***
1	***289***	***3***	***309***
5	***262***	***29***	***291***
10	***262***	***33***	***254***
15	***235***	***56***	***259***
20	***215***	***87***	***215***
25	***217***	***90***	***209***
30	***178***	***133***	***189***
35	***161***	***138***	***153***
40	***149***	***155***	***153***
45	***110***	***192***	***105***
50	***112***	***194***	***107***
55	***81***	***234***	***79***
60	***61***	***245***	***53***
65	***59***	***255***	***53***

5. CONCLUSION

We have demonstrated that our system is able to detect various human gestures like walking, bending, arm stretching using SOTON database and WEIZ MANS ACTION database to test the proposed system. The SOTON database contains the sequence of images from a video with silhouette extracted. WEIZ MANS ACTION contains video for various actions like Walk, Run, Jump, Gallop sideways, Bend, One-hand wave ,Two-hands wave, Jump in place, Jumping Jack, Skip.

6. FUTURE WORK

Although a large number of gesture recognition algorithms and methods have been reported, but it's worth mentioning that gesture recognition is still in its infancy.in future we can combine all the algorithms for various gestures. So, that we can develop a multi-gesture recognition system.

II. REFERENCES

[1]. K. Srijeyanthan, A. Thusyanthan, C. N. Joseph, S. Kokulakumaran, C. Gunasekara and C. Gamage, "Skeletonization in a real-time gesture recognition system," 2010 Fifth International Conference on Information and Automation for Sustainability, Colombo, 2010, pp. 213-218.

[2]. Hasthorpe, J, and Mount, N., The generation of river channel skeletons from binary images using raster thinning algorithms, in the website of Geographical Information Science Research Conference, 2007.

[3]. C. N. Joseph, S. Kokulakumaran, K. Srijeyanthan, A.Thusyanthan, C. Gunasekara and C. D. Gamage, "A feeds," 2010 5th International Conference on Industrial and Information Systems, Mangalore, 2010, pp. 430-435.

[4]. Sen-Ching S. Cheung and Chandrika Kamath, Robust techniques for background subtraction in urban traffic video, In Proceedings of SPIE, The International Society for Optical Engineering, 2004, pp. 881-892.

[5]. C. N. Joseph, S. Kokulakumaran, K. Srijeyanthan, A.Thusyanthan, C. Gunasekara, and C. D. Gamage, Comparison of background subtraction algorithms on video streams, Technical Report, Dept of CSE, University of Moratuwa, Sri Lanka, 2010.

[6]. S. Sclaroff et al., "Tracking, analysis, and recognition of human gestures in video," Eighth International Conference on Document Analysis and Recognition (ICDAR'05),2005, pp. 806-810 Vol. 2.doi: 10.1109/ICDAR.2005.243

[7]. M. Betke, J. Gips, and P. Fleming. The Camera Mouse: Visual tracking of body features to provide computer access for people with severe disabilities. IEEE Trans Neural Sys. Rehab. Eng., 10(1):1–10, 2002.

[8]. J. W. Hsieh, Y. T. Hsu, H. Y. M. Liao and C. C. Chen, "Video-Based Human Movement Analysis and Its Application to Surveillance Systems," in IEEE Transactions on Multimedia, vol. 10, no. 3, pp. 372-384, April 2008.

© 2015-19, IJARCS All Rights Reserved