ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 5, May 2017

AI for Analysis of Human Behavior

[1] Manas Kumar Hati

[1] Department of Electronics and Communication Engineering, Galgotias University, Yamuna Expressway Greater Noida, Uttar Pradesh

[1] mkhatikgp@ieee.org

Abstract: Since human and other animals exhibit Natural Intelligence (NI), for programing machine Artificial Intelligence (AI) is used. The field of AI research is described as the study of smart devices such as a robot, computer etc. that take necessary action according to surrounding environment. Moving object is a prominent area of computer-vision research in the field of video surveillance system. This is not effortless work as persistent manipulation of the subject happens through travel. Across spatial and temporal spaces, each object across motion has different accreditations. In temporal space entity the motion rate varies where as in space entity the size differs. Detection and recognition of individuals is the main focus of this research. Community of people existing video sets are known for the purpose of identifying people and tracking people in crowd scene. The method of context subtraction is used to identify humans. To remove information, the descriptor method of the Oriented Gradient Histogram is applied. Support Vector Machine (SVM) classification system is used to classify the conducted human activity.

Keywords: AI, Background Subtraction, NI, SVM, Threshold.

INTRODUCTION

Artificial Intelligence in Computer Vision is used to learn different ways of interpreting, re-building and comprehending 3D images from its 2D scenes, which are determined by the actual link between the constructs in the particular video. This consists mainly of methods for collecting, interpreting, analyzing and manipulating the digital images. Throughout social gatherings, nation frontier, banks, sports stadium, schools, airports and shopping malls, video editing is common. In the field of computer vision, the question of the identification, monitoring and perception of human activity has acquired significance. Identifying and monitoring these bodies moving objects and identifying their behavior in a video surveillance device is a major task. Currently this is used to control and track various artificial intelligence camera systems[1]. Applications include finding out abnormal activities, giving security using video surveillance, control unit for patients, video sports, traffic management, etc.

This is a well-organized and effective approach used specifically to classify movement-based events and action in the recorded video series. In some cases, further behavior can sometimes display differences due to less video quality, shifting context, contrasting circumstance, different human views, and environmental interference, as well as many changing actors in the environment. Human activity identification is mostly used for human interactions with each other as it provides information about people's identity, their behavior, their personality perception, etc. It is widely used in computer robotics interaction with human beings, which uses depiction of the behavior of many. All of these involve a program that distinguishes different kinds of operation.

RELATED WORK

Human identification and recognition of their behavior is a very active area of research within Artificial Intelligence through computer vision, especially detecting suspicious activity. Flanging Zhang, Lisong Wei, and others suggested the idea for human identification in crowds called Cube surface simulation. They also introduced a novel real-time, effective human detection system in their paper. A novel cube surface model captured by a binocular stereo vision camera solves the human detection problem. The paper first suggested a cube surface model for estimating the 3D background cubes in the

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 5, May 2017

control region, and then developed a shadow-free technique for updating the cube surface model. The paper then introduces a shadow-weighted clustering approach for effective human scanning as well as eliminating false images. But the Human behavior was not understood by this system[2].

Tao and other have created a concept called Fast and Efficient Application of Human Upper-Body Detection and Orientation Estimation in RGB-D Images. Their paper introduces a novel integrated approach for the detection and estimation of human upper-body orientation. This solution does not limit the function form, so that any mix of RGB and depth features in the proposed system will work seamlessly. Estimates of human upper-body recognition and orientation was implemented using a study of wild woods. But that paper failed to detect the executed human action[3].

Researchers suggested a tracking algorithm for head recognition in crowd scenes. Their paper presented a method for detecting multiple heads which can be used in a smart human tracking system. The computational power of such systems is so large that it does not extend to embedded devices. So, their paper suggests a parallel design to increase the approach's performance, so that it can be more useful in embedded platforms. Their algorithm deals only with the technique of head identification, and it also struggles to track long trajectory and people's behavior[4].

Proposed object detection and recognition by Yanan Zhang, Hongyu Wang and Fang Xu is the concept and basis for smart machine robot to consider the natural environment and make smart decisions. Their paper introduces an end-to-end object detection and recognition algorithm based on deep learning, aimed at the precision and real-time efficiency of object detection and service robot recognition in complex scenes. The algorithm has both good accuracy but failed to acknowledge human activities[5].

Hiromasa Taada and others suggested human detection in crowded scenes using prior frames of goal information. In their paper the approach compares at the current frame the target area to related regions. The paper also compares the goal area on current and previous frames. The probability of odd colors at current and previous frames is high, thereby increasing the tracing operation. This only works well in tracking individuals in film, but they do not remember their actions[6].

MyoThida and others suggested two different models for the identification and monitoring of human activity in crowded scenes. First technique called macroscopic modeling technique was used to learn about specific motion patterns in crowded scenes and the second type of technique is microscopic modeling, often based on studying the visual direction of moving events. Their simulation methodology, however, detects the normal movement of crowd scenes but struggles to discern specifics of the actions of individual individuals[7].

PROPOSED METHODOLOGY

This paper includes tracking, detection and behavior recognition of the other people from the existing video sets.

1. Existing Video set:

The existing video sets are captured where various actions are being perform with different formats of video in various background.

2. Preprocessing:

Preprocessing main aim is to reducing noise and for creating device data for extraction of feature

3. Extraction of Frames:

Extraction of frames is most prominent phases in which the input video is converted from the many frames. Length of the video is key factor while calculating value of total frame numbers. Additionally, the transformed frames can be used for recognition of observation, retrieval and operation.

4. Background Subtraction:

Using this form, the foreground is retrieved for encoding. This is an influential step in assessing the people in motion in the picture being shot. The region of interest is human in motion in the identification of

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 5, May 2017

persons in its foreground. Hence, moving individuals can be removed from its background image with the aid of this technique. By subtracting the reference image (original image considered under static background condition) and current image frame it distinguishes person at the moment.

A rigorous process of Background Subtraction must accommodate adjustments in the lighting as well as major long-term shifts in the video environment. This inspection uses the term (x, y, t), here (x, y) represents the video sequence position of the pixel in x and y coordinates, and time dimension (t). A simple way to implement this approach is to find a background image as a reference frame (referred to as' B') and a frame obtained as' C(t)' at t (time interval). By using simple mathematical equations, it is possible to determine the person simply by using image subtraction method for each picture element present in C (t), the number of current images of the picture element is represented by P[C (t)] and decreased by its respective pixel value at certain position of the background scene is referred to as P [B].

The difference image obtained would display some of the components of strength for the place of the picture feature, which are redirected in two frame positions considered for context subtraction following. This method shows good outcome when all the objects in the foreground picture are in motion and all the objects in the background are steady. A threshold (T) is used on this discrepancy picture to make subtraction result easier.

For thresholding mathematical form can be described as:

$$P\{C(t)P(B)\} > T$$

5. Recognition of Human Activity:

This is one of the big innovations which can be practiced in real-time scenarios. So far research is based on basic detection of behavior carried out by people such as driving, walking, hand waving, etc. The main intention of developing HAR framework is to automatically evaluate current incidents, behavior, and from the data collected, to get the necessary context. Figure 3.2 represents the simple HAR-system structure.

6. Descriptor of HOG Feature:

This method uses various methods to minimize the number of resources that illustrate a large set of data. Extraction of the object technique is used to evaluate, segregate in the digitized image, multiple desired features. Once the binary images are created, the next task is to extract the functions. Attributes are the interesting parts of an image that are contained in a compact vector form called feature vectors. The suggested work uses the HOG method to extract the required information from the repetitive dimension of the gradient within the restricted location of the image being considered and is divided into the tiny portion of the connection known as cells and a HOG path is arranged for pixels present within each cell. The classification then reflects on the hysterograms obtained. Extracted attributes, called the HOG function matrix, are used to inform the SVM[8].

Typically, the attributes collected will be the' M matrix' in' 1' where' M' represents the range of features of the HOG. Consequently, the data obtained could be used, categorized and documented thoroughly in categories. As shown in figure 3.4, the extraction of the area of people in motion using a HOG approach. The sample dataset video is taken into account in order to get HOG properties, where the traveling human being is first identified and only area in motion is taken out. Once after this, the HOG method is used to get the characteristics of moving people present and will be illustrated for each of the features collected. To order to obtain the attributes, videos are taken as the input supplied to the gradient calculation block in which direction and magnitude are determined, after which the resulting values are sent to the' Gradient Vote ' block and all HOGs are measured in the final phase.

- **Gradient Computation:**

The magnitude m (x, y) and path (x, y) is calculated for each of the pixels (x, y). Assuming that the luminance value of the pixel to be computed is centered at the coordinates (x, y) and f (x, y). The x-

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 5, May 2017

axis and y-axis gradients are referred to as fx (x, y) and fy (x, y).

Gradient value can be written as:

fx (x; y) = f (x+1; y) f (x 1; y)

fy (x; y) = f (x; y + 1) f (x; y 1)

Direction (x; y) and magnitude m (x; y) is determined as:

m (x, y) = fx (x; y)^2 + fy (x; y)^2

$$(x, y) = \tan \frac{fx\,(x;y)}{fy\,(x;y)}$$

- **Gradient Vote:**

After the magnitude m (x, y) and path (x, y) is collected, gradient vote is determined for all pixels, which in turn gives histogram of orientation. The direction is distributed equally between 0 and 180. The weight of each pixel is given by:

$$j = (n + 0.5)b\,(x; y)$$

Where n indicates a bin where path (x; y) is used, and where the value of b indicates the total number of bins. To eliminate aliasing, two adjacent bins values are incremented. m_a and m_b determined as:

m_a = (1)

m_b = m (x; y)

- **Computation Normalization:**

This technique is applied at the Last stage by combining all the histograms which are a part of a particular block.

- **SVM Classifier:**

It is an organized model of learning. The learning algorithm is used to inspect the classification data that is needed.

IMPLEMENTATION

Using' MATLAB' is applied artificial intelligence for human behavior analysis. The video datasets are taking the initial step in identifying people and HAR (Human Activity Recognition). The databases consist of people taking moments such as standing, sitting, hitting, handshaking, embrace, jumping, and fallback. Implementation of the' HAR' continues with frame extraction. Extraction of the frame, which starts with acquiring knowledge about video file size, memory, quality, resolution, etc. by giving command called 'aviinfo'. Using the' video reader' command video will be read and then number of frames in video will be created using numerates. After all execution of the total frames loop starts until it reaches the end of frames, after that frames are read in video file.

Once the next operation is to convert these frames to image files using the command' frame2im' after obtaining the frames. At the end the converted image is saved. The Frame extraction rate is' 30' frames per second. Another phase is required, because videos cannot be handled directly. Strategy is then used to classify the traveling people. This approach would find a background image in which each picture will be subtracted by this background image in order to obtain foreground images representing the area of humans. Foreground graphic' RGB' is translated to pictures on a' Grey' scale. The resulting '2-D mean filtering' is used to eliminate the noise elements.

When completed with reduction of noise 'Gray scale' images will be transformed to 'Binary images' of zeros and zeros, where binary 0 is used for human absence and binary 1 shows white area of human presence. Hence it is very helpful to remove any moving people and objects present in the creation of the video binary image. During this process the dilation operation is conducted on binary images collected.

Measurements include mostly area that provides actual number of pixels present in the image region, bounding box that represents a small rectangular box in which individual human region resides, centroid that defines the middle pixel of each observed human and many more. It will be identified when doing all the tasks, independently and also group of people. SVM classifier' is made use of to classify human action.

Major step in choosing training sets is to build the classifier. About seven different kinds of training directories are being generated in present work. This

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 4, Issue 5, May 2017

includes the movements such as walking, talking, and handshaking, jumping, hitting, kissing and falling back. The SVM classifier needs training classifier N(N-1)/2 in which N denotes the number of training files. Classification makes use of' One to One Strategy.' When checking, if the result is correct, when a sample is sent to the classifier as input, the category belongs to class A, otherwise belongs to class B.

RESULT

The analysis carried out by video datasets that are captured under all kinds of backgrounds, occlusions and gestures. This is understood using the program MATLAB.

1. **Video Testing:**

A video was a high quality UT-Interaction style dataset. In that video 4 men perform 5 different actions such as hug, slap, stand, walk, and jump. The recorded video has a length of 20 seconds and a buffer size of 2.7 MB. The video's Pixel resolution is 1020 720, with 40 frames per second. The accompanying figures show numerous machine recognized events.

- **Frame 1:**

In frame 1 system recognize first person standing and second & third person walking.

- **Frame 2:**

In frame 2 system recognize first & second person hand shaking and third person standing.

- **Frame 3:**

In frame 3 system recognize first person punching third person and second person standing.

- **Frame 4:**

In frame 4 system recognize first person kicking third person and second person standing.

- **Frame 5:**

In frame 5 system recognize first person standing and second & third person walking.

CONCLUSION

AI plays important role in day today life and human detection by AI is emerging technologies. It can be used for protect the public against acts of criminal violence. In Mumbai blast at the famed 'Taj Hotel', the gunman targeted and occupied the site for about four days. They shot innocent people and made the whole country terrified. If this strategy is applied there, the shooting operation can be considered illegal and the device will immediately notify the security guards. In the field of Artificial Intelligence through computer vision People detection and recognition in crowd is an important and challenging and task. The Future work attempts to recognize other kinds of actions such as chewing, face recognition, walking, mood analysis including the facial expression and other variables that can be achieved through recognizing each particular person's body movement in the film. Future artificial intelligence can be imported into a robot to detect humans and analyze their behavior.

REFERENCES:

[1] A. Intelligence, “Fundamentals of Neural Networks Artificial Intelligence Fundamentals of Neural Networks Artificial Intelligence,” *Fundam. Neural Networks AI Course Lect. 37 – 38, notes, slides*, 2010.

[2] J. Li, F. Zhang, L. Wei, T. Yang, and Z. Li, “Cube surface modeling for human detection in crowd,” 2017, doi: 10.1109/ICME.2017.8019311.

[3] T. Ji, L. Liu, W. Zhu, J. Wei, and S. Wei, “Fast and efficient integration of human upper-body detection and orientation estimation in RGB-D video,” 2017, doi: 10.1109/ICCSN.2017.8230296.

[4] S. Rohatagi, D. Profit, A. Hatch, C. Zhao, J. P. Docherty, and T. S. Peters-Strickland, “Optimization of a digital medicine system in psychiatry,” *J. Clin. Psychiatry*, 2016, doi: 10.4088/JCP.16m10693.

[5] M. Shah and R. Kapdi, “Object detection using deep neural networks,” 2017, doi:

10.1109/ICCONS.2017.8250570.

[6] H. Takada, K. Hotta, and P. Janney, "Human tracking in crowded scenes using target information at previous frames," 2016, doi: 10.1109/ICPR.2016.7899899.

[7] M. Thida, Y. L. Yong, P. Climent-Pérez, H. L. Eng, and P. Remagnino, "A literature review on video analytics of crowded scenes," in *Intelligent Multimedia Surveillance: Current Trends and Research*, 2013.

[8] M. A. Torkamani and D. Lowd, "On robustness and regularization of structural support vector machines," in *31st International Conference on Machine Learning, ICML 2014*, 2014, vol. 3, pp. 1989–1999.