International Journal of Engineering and Technical Research (IJETR)
ISSN: 2321-0869 (O) 2454-4698 (P), Volume-3, Issue-12, December 2015

Survey paper on Human Humiliation Detection based on Human behavior Analysis

Rohan Shiroor, Pranav Ranadive, Nitin Sankpal, Satish Vairagar

***Abstract*— We are describing social issues solution regarding crimes related to humiliation of human. Today we are under the threat of many types of crime scenes among which terrorism and rape cases are the most important and critical issues that we have to worry about. The most common thing among these entire crime scenes is the aggression inside a human being that will lead to more disastrous results. And this paper is about detecting and analyzing the human aggression by using some of the constraints of human behavior. We are not using any of the wearable sensors as it may increase the cost, instead implementing firstly image processing on the various human poses that can be still pictures or sequence of pictures and then secondly applying Bayesian Network to identify the future actions of the human. We will perform visual surveillance task that will then compare the interactions between multiple people and compare the aggression rate of all of the people in that particular scenario and send it back to the Bayesian network to analyze it and take appropriate actions based on the future predictions of the human behavior. These can be applied to the Unmanned Aerial Vehicles equipped with the visual surveillance system to perform all the actions by learning behavior of human in an interaction and apply appropriate action on the most aggressive human being among the people in the scenarios. Unlike the mainstream video surveillance approaches, the proposed method does not rely on background subtraction or dynamic features and thus allows for action recognition in still images.**

***Index Terms*— Bayesian Network, face primitives, human activity recognition, human body joint detection.**

I. INTRODUCTION

Over the years, crimes have been increased in the most of the part of the world till date. The most prominent reason for all of persons interacting in a group or actively participating in the group, these categories are:

1. Actor
2. Receiver

These aggression about something, whether it is humiliation physically or mentally. So to detect such behavior among human is very difficult at an instant of time, but we are demonstrating how to do such types of observation through a machine which will learn the things by stimulating the environment it will observe. There has been growing interest in the most of the machine learning approaches for analyzing human behaviors; such systems typically consist of a low-level or mid-level computer vision system to detect the human behavior through moving objects. The system is particularly concerned with detecting when interactions between people occur and identifying the reactions of interacting people. This paper describes one of the approaches to bring the behavior of the human in contrast using some of the machine learning algorithms that is Bayesian network to detect and analyze human action recognition. Human action recognition aims at automatically telling the activity of a person, that is to identify if someone is walking dancing, or performing other types of activities. The task is challenging due to changes in the appearance of persons, articulation in poses, changing backgrounds, and camera movements. In this work, we concentrate on pose based activity recognition and also make them predictable with the use of Bayesian network. For pose based action recognition we have to target three disjoint problems. We have to

1. Detect a person in the image
2. Recognize the expressed pose, and
3. Assign the pose to a suitable action category.

This system based on the recognizing the behavior of the human in an interaction using visual surveillance, record information on basis of position that is pose primitives and sending recorded information to the machine learning algorithm that is Bayesian network. After we understand the type of machine learning problem we are working with, we can think about the type of data to collect and the types of machine learning algorithms we can try. We have to identify the two different categories.

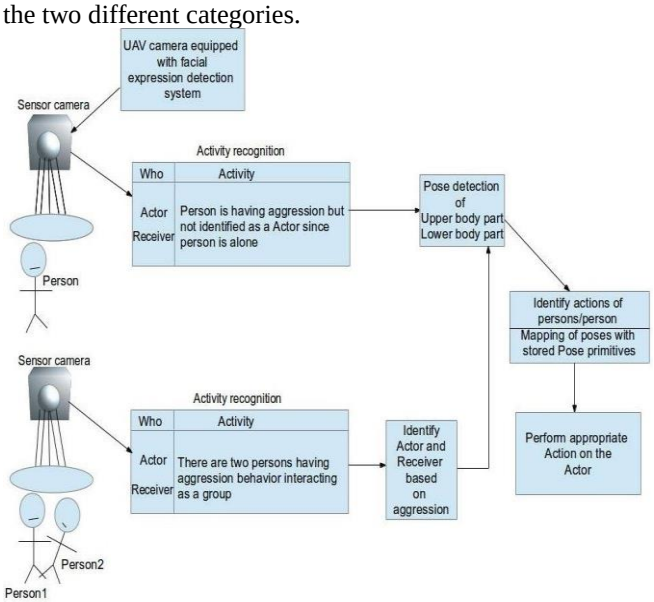

Figure1. Basic setup architectural working of System

Actor is the person who inflicts act of humiliation, responsible for making an aggressive action on the interacting person in the group. A receiver is the person who is feeling the effect of humiliation; person is having somehow less amount of

Rohan Shiroor, B.E. Computer, MAEER's MIT College of Engineering, Pune, India, 8149724921

Nitin Sankpal, B.E. Computer, MAEER's MIT College of Engineering, Pune, India, 9921584661

Satish Vairagar, B.E. Computer, MAEER's MIT College of Engineering, Pune, India

Pranav Ranadive, B.E. Computer, MAEER's MIT College of Engineering, Pune, India, 9730496908.

aggression than that of actors. These primitives can be detected by using the pose or simply positions possessed by a human during defined activity by pointing out their joint and also pose primitives. Aggression may results into humiliation while performing this activity; our system proposes future predictions a particular action by mapping with the stored data.

II. LITERATURE SURVEY

We surveyed on various aspects such as Bayesian networks, visual surveillance and pose primitives. Visual surveillance is an active research topic in image processing. Physical activity can be defined as "Any bodily movement produced by skeletal muscles result in energy expenditure above resting level". There are different ways an algorithm can model a problem based on its interaction with the experience or environment or whatever we want to call the input. Also there were some existing system [4] which proposed activity recognition using Inertial sensing for health-care, and data. Sports Applications which provides the different approaches for activity recognition using inertial sensors. Another system [2] which proposes similar approaches but implementing Bayesian network which increases the efficiency of the system to learn the actions. The only missing thing we observed is that they recognize only fore-arms and hands of a human being which will not give us a perfect behavior of the human. They used the RGB-D sensor which actually senses the interacting object and provide a RGB color distributions the image so that it can differentiate the object form other object. One of the main issues to solve in recognizing human activities is the problem of binding different information sources. There is a large body of work on the analysis of human motion reported in the literature. This theme is addressed in where a multi-modal architecture is used for fusing and interpreting the input from different sources, like voice and gestures. Other authors proposed instead to merge information from object and/or gesture recognition. In an approach is proposed for learning the semantics of object-action relations by observation. Another approach based on Petri nets is proposed for learning human task. A possible way for implementing task recognition is to use probabilistic graphical models. Hidden Markov Models (HMM), Bayesian Networks (BN) and Dynamic Bayesian Networks (DBN) are widely used for speech recognition and bio sequence analysis, but they are used also for task modeling and recognition. Some of the researcher working on the body languages which include ISE labs which aims to design a Cognitive Vision System for human motion and behavior understanding followed by communication of the system results to end-users. Another type of detection done by another researcher who wanted to identify only upper part of the body and performed analysis on that. A survey on Human Activity Recognition[5] using sensors that needed to be wear by the object and so that they can be observed and analyzed with the use of supervised and semi-supervised learning. They are applicable in many of the medical, security, entertainment, and tactical scenarios. The Department of Electrical Engineering, Fu Jen Catholic University, Taiwan conclude with their two-Stage Bayesian Network Method for 3D human pose [1], they have used the most accurate and efficient method of detecting the human pose or their positions in the public as well as private places with the help of human body joints which results from estimation from Monocular Image Sequences. They have performed many experiments on various objects and having efficient and accurate results. The main challenge in structure learning is to develop algorithms that have the computational efficiency of the constraint based algorithms, while relaxing assumptions such as faithfulness, for the underlying distribution. Bayesian networks are a versatile tool of artificial intelligence, as any artificial intelligence in real life must be able to reason probabilistically, in order to cope with uncertainty. They have a wide range of applications; for example, reliability theory, and system security and in bioinformatics, where Bayesian network structure learning techniques are used to locate genome pathways. A Dynamic Bayesian Network, able to represent the task in a probabilistic fashion, is also designed and implemented. With the proposed architecture is possible to infer simple tasks, in a way that is robust to variations in the execution sequence. The experiments presented indicate that the pose of a human already contains sufficient information about the underlying activity.

III. APPLICATIONS

Our proposed system can be applicable to most of the today's common social issues which includes:

Table 1. Various social issues and corresponding applications related to proposed system

Social issues	Censorious	Applicability of proposed system
Ladies Safety	It is the most sensitive and critical issue.	Almost applicable as humiliation can be detected within the group.
Terrorist Attacks	It is the issue facing almost every country in the world.	Applicable if there will be some activity performed by the terrorist in the crowd.
Human public violence	Most consistent problem but may result into a war.	Proposed system may be applicable to the issue if multiple systems will be deployed at the same place so that it may detect multiple activities in the crowd.

IV. PROPOSED SYSTEM

We are supposed to propose the surveillance system which can be deployed in public place or civilian's area where it will take a look over people for the purpose. We have parted all the system in the following way as shown below is the table for providing more information about each part function.

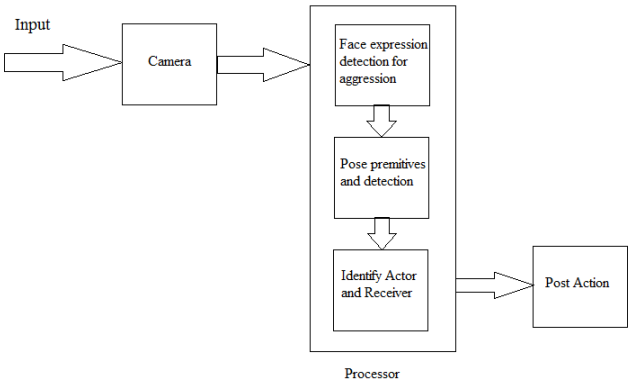

Figure 3. Block diagram for subsystems description of proposed system

Table 2. Proposed system partitioned into some subsystem correspond to some functions

Subsystem	Functions
Facial Expression Recognition	Detect the facial expressions in the public environment in an interaction and identify Actor and Receiver
Pose primitives	Provides information about the poses of the human and make it available to the processor.
Finalize Actor and Receiver	At last processor will process the data and predict the further activity of the human and take appropriate actions on the Actor.

This approach may include some high-tech featured Unmanned Aerial Vehicles that need to be deployed in the surveillance area where we actually need it. Below will be brief discussion about the idea step by step.

1) ***Drone surveillance:***

In May 2014, Mumbai inhabitants witnessed what could easily have been a scene lifted straight from a sci-fi novel; a pizza was home-delivered using an Unmanned Aerial Vehicle (UAV), more popularly known as a drone, from a local pizzeria. This experiment was not amiably met by the local police. A notice was shot off to the allegedly offending outlet which had not taken permission of either the local police or the Air Traffic Control of the Mumbai International Airport raising questions over the legality of commercial usage of UAVs in India.

Drone use has been synonymous with the fiendishly successful military operations run by the American military forces over the Middle East. Critics have lamented the moral and legal grey area in which the US military drone programmer functions. The opinions of these critics were hilariously summarized by famous British-American comedian John Oliver. Despite their arguably reckless military use, the legal and moral debate attached to drone use is easily sidestepped for their hard-hitting tactical benefits. The emergence of this new branch in the military-industrial complex has been responsible for the spillover of military drone technology into civil space, with existing and new players actively exploring the vast possibilities in civilian use. Recent advancements in software technology have allowed a multitude of uses, besides significantly bolstering drone reliability and flying capabilities.

Patrons of the drone industry understand that they are sitting on a gold mine. Their biggest hurdle is navigating the tough regulatory waters for the responsible use and operation of drones for civilian purposes. Civil aviation authorities around the world are finding it hard to regulate civilian drone operations within the existing framework of regulations. The implications that drones will have on law, society and the individual are still being fully understood and raise many safety and privacy concerns. Yet, there is still a lack of consensus on certain key issues like what exactly is a drone? Are they remote-controlled or do they include autonomous vehicles too? Are flying toys also called drones? How will drone regulations be enforced and what will be an appropriate penalty? These are tough questions and are currently being debated in the US, EU and in several other countries. But a License-Raj-style case-by-case approval regime with no specified process for approvals cannot be the solution. By taking into account all the above problems and their solution we have to develop our system and we will look forward to overcome the problem by substituting it with more simple approach but with more complex security system. With above descriptions there is one thing that will describe drones in following way as civilian drones for safety of public premises with the help of surveillance cameras equipped. Hence we can say that civilian drones come with some risks and reward too. Drones will be equipped with too less things so that it should be efficient and reliable to wind over the public area because the height of the drone will be the first and foremost issue that we will take into account. In the end we have to identify Actor and Receiver only and perform particular task on Actor so that violence should be stopped and humiliate person's information should be gathered by local cops. This will be explained in further points. We have to use camera sensors that will help to detect human in the environment other than usual scenarios that will include cars, buses, etc. These cameras help to identify aggression through facial expression recognition algorithms.

2) Aggression identification- facial expressions:

We describe a real time computer vision and machine learning system for modeling and recognizing human behaviors in a visual surveillance task. The system is particularly concerned with detecting when interactions between people occur and identifying the reactions of interacting people. Here we are going to detect the Actor and the Receiver. The Actor is the person who is going to perform any violent or aggressive action on the receiver. The receiver is the one on which any violent or aggressive action is performed. Facial expression identifies the basic human emotions .It helps to tell the person by watching the image whether the person is actually telling the truth about what he is claiming or not. Human emotions detection is used in human computer interaction, military, law-enforcements etc. This system based on recognizing the behavior of the human in an interaction using visual surveillance, record information on basis of Facial movements and sending recorded information to the machine learning algorithm.

System is used for further processing and to recognize the exact expression. FACS decomposes facial expressions in terms of 46 component movements, which roughly correspond to the individual facial muscles. Using FACS, practically all facial muscle movements can be accurately described in terms of Action Units or Facial Action Units, which appear to be the smallest possible changing units in a face. After the tracking the FAUs in the wireframe grid are detected before employing them to produce one of the six basic facial expressions using a set of rules that maps them to facial expression. In the case of FAU-based facial expression recognition, for every FAU, the database is clus- tered into two different classes. The first class, represents the presence of the FAU under examination at the grid being processed, while the second one, represents its absence. Facial expressions can be described as combinations of FAUs.

3) Pose detection through primitives

For pose based action recognition, a reliable representation and recognition of individual poses is crucial. Most difficulties in pose matching arise from cluttered background and pose articulations. Often, background objects are falsely recognized as limbs or parts of a pose. We recognize poses by matching them to a set of learned pose primitives.

We have to consider the social signal processing as deploying it at the civilian area Social Signal Processing aims at developing theories and algorithms that codify how human beings behave while involved in social interactions, putting together perspectives from sociology, psychology, and computer science. Here, the main tools for the analysis are the social signals, i.e., temporal co-occurrences of social or behavioral cues that can be basically defined as a set of temporally sequenced changes in neuromuscular, neuro-cognitive, and neurophysiological activity. It is a very important and challenging problem to track and understand the behavior of agents through videos taken by various cameras. The primary technique employed is computer vision. Vision-based activity recognition has found many applications such as human-computer interaction, user interface design, robot learning, and surveillance, among others. In vision-based activity recognition, a great deal of work has been done. Researchers have attempted a number of methods such as Hidden Markov models, etc., under different modalities such as single camera, stereo, and infrared. In addition, researchers have considered multiple aspects on this topic, including single pedestrian tracking, group tracking, and detecting dropped objects.

Image features, such as edge, color, and silhouette, are observations of a pose. The extraction of image features constitutes evidence nodes of the articulated human model for the inference in the Bayesian network. Single human feature is not enough to inference 3D human position since different 3D poses can exhibit similar 2D observations in the images. Therefore, we devise 4 kinds of features in the proposed method: human silhouette, normalized center of human body, spatial distribution of skin color, and corners of human body.

So here was the brief description of what we will do in this part. It is a short description but very important part of the proposed system. Interactions should be identified before reaching to this part as violence scene requires interaction between two or more people. There may be many Actors and one Receiver but it can be vice versa too.

V. CONCLUSION

Human behavior understanding is a complex and very difficult problem, which is still far from being solved in a way suitable for anticipatory interfaces and human computing application domain. In the past two decades, there has been significant progress in some parts of the field like face recognition and video surveillance (mostly driven by security applications), while in the other parts of the field like in non-basic affective states recognition and multimodal multi-aspect context-sensing at least the first tentative attempts have been proposed. We tried to describe the all the things required to implement Human Humiliation detection.

Although the research in these different parts of the field is still detached, and although there remain significant scientific and technical issues to be addressed, we are optimistic about the future progress in the field. The main reason is that anticipatory interfaces and their applications are likely to become the single most widespread research topic of AI research communities. We are looking forward to implement these systems on the UAV's so that it can make the issue mentioned in the report to be solved and improve by time. It may be a combination of various concepts described or on a single efficient system. There may be some limitations among the sensors used as it will be mobile and sometimes may not absorb the sufficient data from the environment but still this system provides efficient results for getting the appropriate results.

ACKNOWLEDGMENT

Author thanks to the mentor prof. Nagesh Jadhav to contribute their expertise views for this paper and guidance for further developments, we want to give special thanks to experts who contributed indirectly or directly to develop this paper. Last but not the least; we are thankful to MIT College of Engineering for their supportive approach.

REFERENCES

[1] Volume 2010, Article ID 761460 Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park. "A Two-Stage Bayesian Network Method for 3D Human Pose Estimation from Monocular Image Sequences"

[2] Dec ,2011, "A Bayesian Approach for Task Recognition and Future Human Activity Prediction", Vito Magnanimo , Matteo Saverianot , Silvia Rossi and Dongheui Leet

[3] Oct,2011,"Human Behavior Analysis in Video Surveillance: a Social Signal Processing Perspective", Marco Cristani1;R.Raghavendra1 Alessio Del Bue1 Vittorio 1Istituto Italiano di Tecnologia (IIT), Genova, Italy Dipartimento di Informatica, University of Verona, Italy

[4] "Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey" Akin Avci, Stephan Bosch, Mihai Marin-Perianu, Raluca Marin-Perianu, Paul Havinga University of Twente, The Netherlands

[5] VOL. 15, NO. 3, 2013,IEEE COMMUNICATIONS SURVEYS & TUTORIALS," A Survey on Human Activity Recognition using Wearable Sensors O´ scar D. Lara and Miguel A. Labrador"

[6] 2008, "Video Behavior Profiling for Anomaly Detection", Tao Xiang1 and Shaogang Gong Department of Computer Science

[7] 2008, "Recognizing Emotion From Facial Expressions: Psychological and Neurological Mechanisms", Ralph Adolphs, University of Iowa College of Medicine

[8] "Facial Expression Recognition in Image sequences using Geometric Deformation Features and Support Vector Machines", IEEE Transactions on Image Processing,Vol.16,No.1

Rohan Shiroor pursuing Bachelor of Engineering in computer stream at MAEER's MIT College of Engineering, pune have achieved excellence in the Sixth Sense Robotics and Advance Robotics. Also an active member of IEEE Pune Region, contributed to open source projects on Github community.

Nitin Sankpal pursuing Bachelor of Engineering in computer stream at MAEER's MIT College of Engineering, pune have achieved excellence in the Sixth Sense Robotics and Advance Robotics. Also an active member of IEEE Pune Region, contributed to open source projects on Github community.

Satish Vairagar pursuing Bachelor of Engineering in computer stream at MAEER's MIT College of Engineering, pune have achieved excellence in the Sixth Sense Robotics and Advance Robotics. Also an active member of IEEE Pune Region, contributed to open source projects on Github community.

Pranav Ranadive pursuing Bachelor of Engineering in computer stream at MAEER's MIT College of Engineering, pune, India, have achieved excellence in the Sixth Sense Robotics and Advance Robotics. Also an active member of IEEE Pune Region, contributed to open source projects on Github community