International Journal of Research in Advent Technology (E-ISSN: 2321-9637) Special Issue
National Conference "NCPCI-2016", 19 March 2016
Available online at www.ijrat.org

Survey on Selection of Features Used for Anomaly Detection

Sharadhi Rao[1], Aneeha Patwardhan[2], Maushumee Vyavahare[3], Prof. Kishanprasad Gunale[4]
Department of Electronics and Telecommunication, MIT College of Engineering, Savitribai Phule Pune University, Pune, India[1,2,3,4]
Email: sharadhi.s.rao@gmail.com[1] aneeha@gmail.com[2] maushumee.vyavahare@gmail.com[3] kishanprasad.gunale@mitcoe.edu.in[4]

Abstract – In this paper, the features that are most commonly used in anomaly detection have been reviewed and described. Modeling human behavior and activity patterns for recognition or detection of abnormal event has attracted significant interest for research in recent years. Diverse methods are plenty for building intelligent vision systems aimed at scene understanding and making correct inference from the observed dynamics of moving targets. Most applications are in surveillance, video content retrieval, and human–computer interfaces. These can be used to develop surveillance-based systems, which can be used in hospitals, old-age homes, for fall detection, security, purposes, etc**.**

***Keywords*— Depth, Feature, Classifier, Motion, Threshold, Silhouette, Skeleton Mapping**

1. INTRODUCTION

Learning by machine and understanding human actions is a complex, diverse, and challenging area that has received much attention within the past years. Human action detection, motion tracking, scene modeling, and behavioral understanding (human activity recognition and discovery of activity patterns) have garnered a lot of attention in the computer-vision and machine-learning communities. Applications have been in video surveillance, human–computer interfaces, and multimedia semantic annotation and indexing. Intelligent visual surveillance has got more research attention and funding due to increased global security concerns and an ever increasing need for effective monitoring of public places such as airports, railway stations, shopping malls, crowded sports arenas, military installations, etc., or for use in smart healthcare facilities such as daily activity monitoring and fall detection in old people's homes. Often, the objective is to detect, recognize, or learn interesting events, which contextually may be defined as "suspicious event", "irregular event", "uncommon behavior", "unusual activity/event/behavior", "abnormal event", "anomaly", etc.[1]

According to the World Health Organization, approximately 28-35% of people aged 65 and over fall each year increasing to 32-42% for those over 70 years of age. The frequency of falls increases with age and frailty level. Falls increase with age-related biological changes linearly, which leads to a high incidence of falls and fall related injuries in the ageing societies. If preventive measures are not taken in the immediate future, the number of injuries caused by falls is projected to be a 100% higher in 2030. In this context, assistive devices that could help to reduce this major health problem are a social necessity. Hence, fall detectors are being actively investigated.

A fall detection system can be defined as a device which assists in generating an alert when a fall has occurred. In real-life scenarios, they have the potential to alleviate some of the adverse impacts of a fall. Specifically, fall detectors can have a direct impact on the reduction in the fear of falling and the rapid provision of assistance after a fall. Fear of falling has been shown to be associated with negative consequences such as avoidance of activities, less physical activity, anxiety, decreased social contact and lower life quality. [2] Studies conducted with community alarm users who had experienced a fall in the previous six months showed that they felt more confident and independent, and considered that the detector improved their safety. One of the conclusions of the study was that the fear of falling is likely to be reduced by user perception of the reliability and accuracy of the fall detector.

Falls in the patients and in the elderly is a major public health problem because of their frequency and resulting medical consequences. New smart assistive technologies make it possible to provide more security.

A smart system can automatically monitor activities in the range of camera enabling early warnings.

Typically, a video-based sensor uses video surveillance to monitor the user's condition as well as applies digital image processing applied to the real-time video-captured imaging in order to detect whether or not a falling event has occurred. In such video based sensors, digital image processing involves preprocessing of the captured video frames, extraction of suitable features from the real time surveillance video and anomaly detection by using an appropriate classifier. The feature extraction in image processing plays an important role as based on this information, which is given as input to the classifier, anomaly is detected. As shown in figure 1, the movement of the person in the video is captured by the form and position of the shape representation in the panel below. A reliable system and efficient surveillance video system needs to be robust. Therefore, the correct choice of camera, the position of the camera, and appropriate video compression method are important aspects to be considered. [5]. The advantage of the video surveillance system is that it provides a secure and quick intervention for senior citizens, and the video-tape images before falls give important information to give a better understanding of the origins of the falls [6].

In an intelligent video surveillance system, the video will be continuously captured and the algorithm will detect the falls or anomaly and will take a specified action against the anomaly detected. For this, it is necessary to capture or extract appropriate features and then feed those features to the classifier. Finally, depending on the output of the classifier, system will automatically notify the concerned authority if anomaly is detected. As stated before, this paper consists of certain features (such as depth information, skeleton mapping, motion detection etc.) which can be extracted from the video inputs. The classifier in the system will compare these extracted features to the reference values and will take further action as specified in the algorithm.

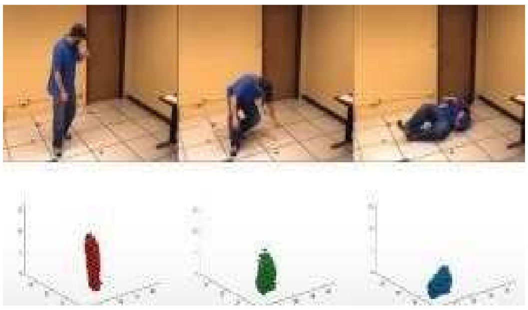

Fig.1: Falling Events and their Representations

2. FEATURESUSEDINANOMALYDETECTION

The features that are commonly used in image processing for anomaly detection are reviewed here.

2.1 HUMAN SHAPE CHANGE

The moving person is first extracted from the image with a background subtraction method. However this is done by taking into account the problems of shadows, highlights and high image compression. Using moments, the person is then approximated by an ellipse defined by its centre ($\bar{x}$, $\bar{y}$), its orientation θ and the lengths a, b of major and minor semi-axes. The approximated ellipse gives information about the shape and orientation of the human in the image. The following figure shows examples of ellipse approximation and subtraction of background.

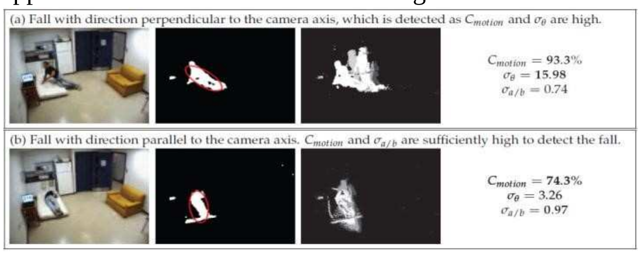

Fig 2: Human shape change

Two features are computed for 1second to analyse the human shape change:

1. The orientation standard deviation σ_θ of the ellipse: If a person falls perpendicular to the camera optical axis, then the orientation will change significantly and σ_θ will be high. If the person just walks, σ_θ will be low.

2. The a/b ratio standard deviation $\sigma_{a/b}$ of the ellipse: If a person falls parallel to the camera optical axis, then the ratio will change and $\sigma_{a/b}$ will be high. If the person just walks, $\sigma_{a/b}$ will be low.

As we want to quantify the motion of the person, we compute a coefficient Cmotion based on the motion history within the blob representing the person using:

$$C_{motion} = \frac{\sum_{Pixel(x,y)\in blob} H_{\tau}(x,y,t)}{\#\, pixels \in blob}$$

Eq. (1)

Here blob is blob of the person extracted using background subtraction, and Hτ is the Motion History Image. Only the largest blob is considered.

This coefficient is then scaled to a percentage of motion between 0%, no motion, and 100%, full motion.

The duration of a fall is extremely short, typically less than a second. So, the Motion History Image by accumulation of motion during 500ms is computed. Motion is taken to be a possible fall if the coefficient Cmotion is greater than 65%.

2.2 DEPTH

A depth map is an image or image channel that contains information relating to the distance of the surfaces of scene objects from a viewpoint. The term is related to and may be analogous to depthbuffer, Z-buffer, Z-buffering and Z-depth. The "Z" in these latter terms relates to a convention that the central axis of view of a camera is in the direction of the camera's Z axis, and not to the absolute Z axis of a scene.

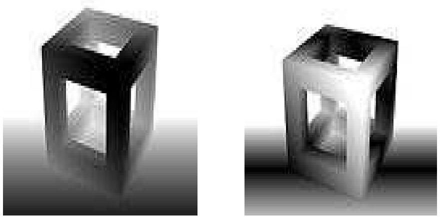

Fig. 3: Depth Maps

Two different depth maps are shown here. The first depth map shows luminance in proportion to the distance from the camera. Nearer surfaces are darker; further surfaces are lighter. The second depth map shows luminance in relation to the distances from a nominal focal plane. Surfaces closer to the focal plane are darker; surfaces further from the focal plane are lighter.

2.2.1 *Uses of depth maps*

- Simulating the effect of uniformly dense semi-transparent media within a scene (fog, smoke or large volumes of water).
- Simulating shallow depths of field - where some parts of a scene appear to be out of focus. Depth maps are used to blur an image selectively to varying degrees.
- Z-buffering and z-culling, techniques, which can be used to make the rendering of 3D scenes more efficient. They can be used to identify objects hidden from view and may therefore be ignored. In real time applications such as computer games, where a fast succession of completed renders must be available in time to be displayed at a regular and fixed rate, this is important.

2.2.2 *Limitations*

- Single channel depth maps record the first surface seen, and so do not display information about those surfaces seen or refracted through transparent objects, or reflected in mirrors. This limits their accuracy in simulating depth of field or fog effects.
- Single channel depth maps cannot convey multiple distances where they occur within the view of a single pixel. This may occur where more than one object occupies the location of that pixel. This could be the case - for example - with models featuring hair, fur, or grass. More generally, edges of objects may be ambiguously described where they partially cover a pixel.

2.3 SKELETON MAPPING

The aim of the skeletonization is to extract a region-based shape feature representing the general form of an object.

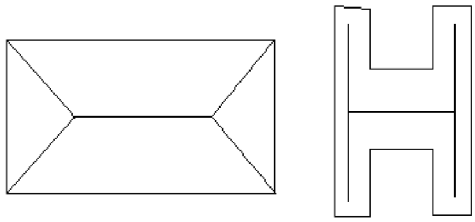

Figure 4: Skeleton Mapping

2.4The skeleton and its properties The notion skeleton was introduced by H. Blum because of the Medial Axis Transform (MAT) or Symmetry Axis Transform (SAT). The MAT determines the closest boundary point(s) for each point is in an object. An inner point belongs to the skeleton if it has at least two closest boundary points.

A definition of the skeleton given by the prairie-fire analogy states that the boundary of an object is set on fire and the skeleton is the loci where the fire fronts meet and quench each other. The third approach provides a formal definition: the skeleton is the locus of the centres of all maximal inscribed hyper-spheres (i.e., discs and balls in 2D and 3D, respectively).

2.3.1 . *Skeletonization techniques*

Skeletonization (i.e., skeleton extraction from a digital binary picture) provides region-based shape features. It is a common pre-processing operation in raster-to-vector conversion or in pattern recognition.

There are three major skeletonization techniques:

- detecting ridges in distance map of the boundary points,
- calculating the Voronoi diagram generated by the boundary points, and the layer-by-layer erosion called thinning.

Only an approximation to the "true skeleton" can be extracted in digital spaces. There are two requirements that have to be satisfied:

- topological (to retain the topology of the original object),
- geometrical (forcing the "skeleton" being in the middle of the object and invariance under the most important geometrical transformation including translation, rotation, and scaling)

2.4 SILHOUETTE

The contour of human region is defined as the human silhouette.

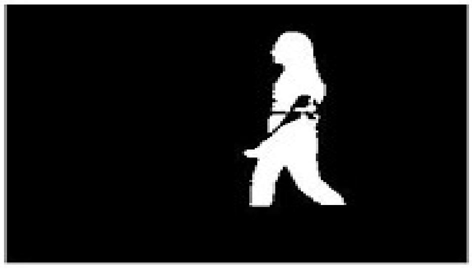

Figure 5: Silhouette There are two sub-systems to calculate the silhouette. The first sub-system is a simple background subtraction method which can be applied in the fixed network camera system. With a pre-defined background image without people and a threshold method, the pixels of foreground can be viewed as the human region.

In the second system which is the pan-tilt camera system or a moving platform, such as robots a human region segmentation algorithm based on normalized-cut energy minimization has been proposed. In traditional normalized-cut algorithm, the image segmentation problem could be formulated as a graph optimization problem. Each pixel can be viewed as a graph node and the linked edges can be modelled by the colour similarity and spatial relationship. This weighted graph can be represented as G = (V,E,W), where V is the collection of pixel nodes, E represents the connection edges with neighbour nodes, and W recorded the weights between nodes which can be calculated by pixel similarity.

3. CONCLUSIONS

Apart from the above given features, there are many others which can be extracted for anomaly detection. Depending on the anomaly, the suitability and accuracy of each method varies. The features thus extracted can be classified using a number of appropriate classifiers which will be useful in determining the occurrence of anomaly.

4. ACKNOWLEDGMENT

The authors would like to attribute their success in this venture to their guide, Prof.K.G.Gunale and his endeavour for perfection, innovation and dynamism and thankful to the Head of Department (E&TC), Prof.Dr.V.V. Shete who helped in every possible way.

REFERENCES

[1] Oluwatoyin P.Popoola, Member, IEEE, and Kejun Wang, "Video-Based Abnormal Human Behaviour Recognition—A Review" in IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, IEEE 2011

[2] Zhen-Peng Bian, Student Member, IEEE, JunhuiHou, Student Member, IEEE, Lap-Pui Chau, Senior Member, IEEE, and Nadia Magnenat-Thalmann, "Fall Detection Based on Body Part Tracking Using a Depth Camera" in IEEE Journal Of Biomedical And Health Informatics, IEEE2013

[3] C Rougier, J Meunier, A St-Arnaud, "Fall Detection From Human Shape And Motion History Using Video Surveillance" in AINAW 2007

[4] Bewley, A., Guizilini, V., Ramos, F., &Upcroft, B. 2014,"Online Self- Supervised Multi-Instance Segmentation of Dynamic Objects"

[5] Jay Chen, KarricKwong, Dennis Chang, Jerry Luk, RuzenaBajcsy, "Wearable Sensors for Reliable Fall Detection " in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference. Shanghai, China, September 2005.

[6] Szu-Hao Huang and Ying-Cheng Pan National Tsing Hua University, Hsinchu, Taiwan, "Learning-based Human Fall Detection using RGB-D cameras" in MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN.