Computer Vision-based Human Activity Recognition for Elderly Home Care

L. Aneesh Euprazia, K.K. Thyagharajan

Abstract: Video-based monitoring of elderly people at home receives more attention in recent days. In this paper, we propose a novel approach to develop smart monitoring system for elderly people using computer vision techniques. Gaussian Mixture Model (GMM) based algorithm is used for background and foreground separation inorder to track the activities of human object. The minimum bounding box of the human object is traced and features like major axis length, minor axis length and orientation angle are extracted. The proposed approach is evaluated on the video sequences of fall dataset.

Keywords: aging population, bounding box, Gaussian Mixture Model, smart monitoring.

I. INTRODUCTION

Elderly people living alone becomes unavoidable now-a-days. These senior people require continuous monitoring to recognize abnormal activities due to health problems. Appointing a nurse for them or hospitalize them for 24hours is not practically possible. But they need some form of assisted living. At the same time they want to be independent. This requirement creates a demand for smart monitoring of key activities which depend on their health conditions. Regular activities of elderly people give a clue about their health conditions. Researchers are always investigating for technology assisted smart home designs to ensure that the elderly live comfortably in their homes. The monitoring systems available in the market are mostly wearable sensors based systems. The outputs of these sensors are connected to a cloud system for further analysis. These devices will also have alarm or emergency buttons to call for help. But this method is useless if the person is unconscious due to abnormal activity like falling down or generally if the elderly people fail to remember to wear them. In addition to this, wearable sensors are available only to certain activities and they are intrusive and requires adherence by the elderly people. The foremost benefit of using wearable sensors is the size of the data output from these sensors is very small.

The alternate solution for wearable sensors is using video sensors for life care and health care of senior persons. Video recording provides a non-intrusive and passive solution for automated monitoring of daily activities. But video cameras are sensitive to lighting variations and the amount of data produced is huge. The latest communication and network technologies make the video data possible to transmit and enable remote safety monitoring of the activities. Analyzing video data requires state-of-the-art computer vision-based approaches and algorithms. The main advantages of computer vision-based human activity analysis are: 1. the system does not require a person to wear anything for the purpose of monitoring. 2. These algorithms provide more information on the behavior of a person. So, it is possible to classify clinically relevant activities accurately and ensure safety of world's aging population. 3. Vision-based home health care monitoring is useful for monitoring medication, meal time and how many meals the person eats a day, sleep duration, sitting, standing and walking duration of the elderly people and provides necessary assistance for daily living of elderly persons. 4. Computer vision based systems can also identify the health decline and provides real-time alerts to the caregivers for earlier medical intervention which will lead to better clinical outcomes.

The main disadvantage of using computer vision methods for human activity recognition is the violation of privacy, handling large sized data, and they are also subjected to vary if illumination varies. But these problems can be solved by using depth camera or thermal camera or both for multimodal data analysis. So, the prime objective of this research work is to progress towards a smart senior home that uses computer vision algorithms to help the elderly persons to meet the challenges of worldwide aging population. The second objective is to develop a Human Action Recognition (HAR) system with reduced cost, complexity and improved recognition accuracy. We believe that human activity recognition and analysis will play a prominent role in the health care of senior people.

I. LITERATURE SURVEY

The computer vision-based systems developed for elderly home care aim to recognize the following abnormal activities: forward fall, vomit, faint, backward fall, chest pain, head-ache, walking speed and gait patterns. Since computer vision techniques require videos captured by a video camera for analysis, privacy becomes a main concern. To ensure privacy of elderly people, the activities are recorded as binary silhouettes. This method is required when RGB video cameras are used. But, if depth sensors are used to capture activity video privacy is protected. In depth cameras, the pixel values in the frame indicate the distance between the object and the camera. Since these cameras do not capture appearance details and the image captured does not depend on the colour, texture or illumination and contains only 3D information about the structure of the objects they preserve privacy. So, depth image will

Revised Manuscript Received on November 27, 2019.
∗ Correspondence Author
  L. Aneesh Euprazia[1] . Full Time Research Scholar, RMD Engineering College, Mail: anueuprazia@gmail.com,
  K.K. Thyagharajan[2], Dean (Academic), RMD Engineering College[1,2]
  Mail: kkthyagharajan@gmail.com,

provide only shape data. The other way of protecting privacy is using thermal cameras.

In these cameras, the pixel value in the images will be corresponding to the temperature of human objects. The temperature of the human body is higher than the surrounding environment and the image produced do not suffer due to low lighting. These images are also invariant to colour of the object. Thermal videos occupy less number of bytes in storage medium compared to RGB videos. The following challenges are met in activity recognition when computer vision techniques are used.

A video clip may be falsely classified as walking if the elderly person is sitting and the nurse walks around the room. Similarity in postures of different activities may create confusion in classifying the events. For example a forward fall and vomiting may be classified wrongly. Long term analysis of continuous activities requires high computational and storage resources. Real-time analysis and remote monitoring requires latest high speed communication networks and high speed computers.

In elderly monitoring system, blind spot is life threatening. So in [1] an Omni-directional vision camera is used. This camera provides a 360^0 panoramic image of the senior's room. These images will have distortions when they are transformed from sensor plane to image plane (bird view technique) for further processing. Prasad calyam et al [2] and Monish Parajuli et al [3] used the depth camera provided in Microsoft Kinect to track the gait parameters such as walking speed, stride length and stride time. Gait patterns of fallers and non-fallers vary. The Kinect has an infrared laser projector combined with a monochrome CMOS sensor. This captures video in 3D irrespective of lighting conditions. This device also has a RGB camera. Falls are one of the major problems for the people aged above 65 because it leads to morbidity and mortality [4]. Zelun luo et al. [5] fed multimodal information (they fused depth modality and thermal modality images) to a convolutional neural network to detect a wide range of activities. Combining two modalities on the pixel levels by concatenating the images channel wise is called early fusion. If the unimodal features are reduced to concept scores and then integrated then it is called as late fusion. Generally fused features provide better performance. If the position of the camera changes, then the the features extracted will also change. So, more than one camera is used to get view invariance features. Shape descriptors are widely used in action recognition. The human shape will change rapidly when a person falls. But the shape will change slowly during normal activities. The shape of the human body during activities is represented by bounding box. But the characteristics of this bounding box will change if the camera angle changes. Human shape can be represented using ellipse shape accurately. Jia-luen et al [6] represented human body shape with three points which are the centroids of head, body and leg regions. Any change in action will have different changes vision the positions of these points. In paper [7] R-transform is used as a shape descriptor, which extracts human silhouette features by using radon transform.

The data set (SDU Fall) created by Ma X et al. [12] provides color, depth and skeleton channels for six action classes such as sitting, walking, bending, falling, lying and squatting. They used an extreme learning machine for classification of the curvature scale space features extracted from action videos. They used a low cost Microsoft's Kinect depth camera for creating this dataset and this is widely used for fall detection. Bingbing Ni et al. [13] created RGB depth Human Daily Activities (RGB-D HuDaAct) dataset which has high intra-class variations. Jiang Wang et al. [14] created a multi view action 3D dataset named North Western – UCLA using three cameras kept at different angles. So, this dataset can be used for cross view action recognition. They also developed a spatio temporal AND-OR graph for recognizing actions. Hanbo Wu et al. [15] combined the above three datasets with MSR Daily Activity 3D and Cornell activity data set (CAD) and created a new dataset with 4528 samples of seven major categories. These actions have 46 subcategories and those had been performed by 74 persons. The MSR dataset was created by capturing videos using single camera fixed only at one angle. It contains 16 activities, whereas, CAD is an imbalanced dataset with 12 categories of activities performed by four persons. It also has skeleton information along with RGB-D images.

II. PROPOSED METHOD

Human activity recognition basically requires the human object to be separated from the background of the image. In this work we use Gaussian Mixture Model for this purpose. Here 'k' numbers of Gaussians are used to describe the distribution of background and foreground pixels in each frame. This can also be used to identify 'k' number of objects in the video. If 'x' is the gray level intensity values of the pixels in the image, μ is the average intensity value and σ is the variance of the pixels then the Gaussian G(x ; μ, σ) is given by

$$G(x; \mu, \sigma) = \frac{1}{\sqrt{2\Pi\sigma}} \, e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \qquad (1)$$

If 'k' number of Gaussians are defined with different μ and σ values, each Gaussian can be assigned a weight. This forms Gaussian Mixture Model (GMM). The intensity of every pixel at any time 't' is represented by GMM.

$$\text{GMM(x)} = \sum_{k=1}^{K} \left(w_k * G\left(x; \mu_k, \sigma_k\right) \right) \qquad (2)$$

Here w_k is the weight, μ_k is the mean value and σ_k is the standard deviation of the kth Gaussian. The sum of all the weights should be equal to 1.

Algorithm for background separation

Input: Pixel x^t

Output: Assigned Gaussian model

Step 1: The pixel x^t of the foreground object with fixed weight value 'w_k' will be assigned to k^{th} Gaussian if it satisfies the following matching condition.

$$M_k(x^t) = \begin{cases} 1 & if \ \dfrac{x^t - \mu_k^t}{\sigma_k^t} < 2.5 \\ 0 & otherwise \end{cases} \qquad (3)$$

Step 2 : If the pixel falls in the kth Gaussian, update the values w_k^t, μ_k^t and σ_k^t of the kth Gaussian using the following equations

Updating w:

$$w_k^t = (1-\alpha) \ w_k^{t-1} + \alpha \qquad (4)$$

Here α is called the learning rate which is constant throughout the algorithm. Its value can be fixed to 0.01.

Updating μ:

$$\mu_k^t = (1- \rho_k) * \mu_k^{t-1} + \rho_k x^t \qquad (5)$$
$$\rho_k = \alpha * G(x^t; \mu_k^{t-1} + \sigma_k^{t-1}) \qquad (6)$$

updating σ :

$$(\sigma_k^t)^2 = (1- \rho_k) * (\sigma_k^{t-1})^2 + \rho_k (x^t - \mu_k^t)^2 \qquad (7)$$

Step 3: If x^t does not fall in any of the K Gaussians, create a new Gaussian with low weight, $\mu = x^t$ and large value of σ. Add the created Gaussian as one of the K Gaussians and remove the Gaussian with least weight.

Step 4: Sort the K Gaussians in descending order based on w/σ after assigning all pixels to any one of the K Gaussians.

Step 5: Fix a threshold T for weight w. For example T may be fixed as 0.7.
The B Gaussians which have $w_k > T$ are the background models.

Step 6: The pixels which are in the remaining Gaussian models represent the object.

The human object is identified in each frame using GMM, which is actually tracking the object in the video as given in figure 1. The minimum bounding box which contains the human object is estimated and it may be used for detecting the human activities. The features such as aspect ratio or width of the person to his height are extracted and used for fall detection in [9] and [10]. Yie-Tarng Chen et al. [11] used elliptical shape to represent the human object. They estimated the ratio of the major to minor axis of the ellipse and its orientation to detect the change in the human shape.

We computed the height to width ratio of the rectangle bounding box, major to minor axis ratio of elliptical bounding box and the orientation angle for activities and recorded in Table-I. The orientation angle is the angle between the major axis and the x axis. We used these features to classify the life threat actions for the elderly people.

Multiple cameras fall datset has 24 video sequences captured by 8 IP cameras for different scenarios. It has action clip of fall and confounding events.

Fig.1 and Fig.2 are the sample frames taken from the video sequences of multiple cameras fall dataset.
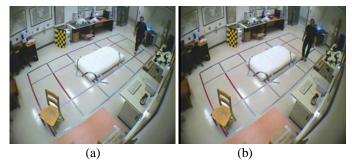

(a)   (b)

Fig.1. Standing

(a)   (b)

Fig.2. Falling

Table- I Activities of Person 1

Activity	Height /width	Major axis / Minor axis	Orientation angle
Standing Fig.1. (a)	2.223	2.44	79.16
Walking Fig.1. (b)	2.043	2.66	72.13
Falling Fig.2. (a)	1.948	1.90	81.08
Falling Fig.2. (b)	1.4848	1.60	87.08

301

Fig.3 and Fig.4 are the sample frames taken from the video sequences of wecare fall dataset.

(a)   (b)

Fig.3. Standing

(a)   (b)

Fig.4. Falling

Table-II Activities of Person 2

Activity	Height /width	Major axis / Minor axis	Orientation angle
Standing Fig.3. (a)	3.6	3.70	-87.01
Standing Fig.3. (b)	3.22	3.64	-86.63
Falling Fig.4. (a)	2.07	2.4	-73.10
Falling Fig.4. (b)	0.419	2.57	-10.56

From the above tables, we show that the height to width of a person is high when he is standing normally and understand that when a person is slowly falling, this ratio values are decreasing. The major axis/ minor axis value decreases for falling activity. The orientation angle is less for normal activity and it increases for abnormal activity.

IV CONCLUSION

In this paper, we proposed a novel approach for smart monitoring of elderly people in three steps: (1) background separation using GMM (2) tracing minimum bounding box (3) extracting minor and major axis length, orientation angle. The features extracted from the minimum bounding box are discriminative. Normal activities (sleeping, standing) and abnormal activities (falling) are recognized. The proposed approach is capable of tracking the activities accurately.

REFERENCES

1. Yun Xiang, Yi-ping Tang, Bao-qing Ma1 Hang-chen Yan, Jun Jiang, Xu-yuan Tian, "Remote Safety Monitoring for Elderly Persons Based on Omni-Vision Analysis" , 2015
2. Prasad Calyam, Isa Jahnke, Anup Mishra, Ronny Bazan Antequera, Dmitrii Chemodanov,and Marjorie Skubic, " Living Lab for Sensor-Based Health Assessment and Physical Therapy", IEEE CLOUD COMPUTING, 2017
3. Monish Parajuli, Dat Tran, Wanli Ma and Dharmendra Sharma, " Senior Health Monitoring Using Kinect", IEEE,2012
4. Jeffrey M. Hausdorff, Dean A. Rios, Helen K. Edelberg, " Gait Variability and Fall Risk in Community-Living Older Adults: A 1-Year Prospective Study", Arch Phys Med Rehabil, 2001, Vol 82, August
5. Zelun Luo,Jun-Ting Hsieh, Niranjan Balachandar,Serena Yeung,Guido Pusiol,Jay Luxenberg,Grace Li,Li-Jia Li,N. Lance Downing, Arnold Milstein, Li Fei-Fei, " Computer Vision-Based Descriptive Analytics of Seniors Daily Activities for Long-Term Health Monitoring", Proceedings of Machine Learning Research 85:1-18, 2018
6. Jia-Luen Chua , Yoong Choon Chang ,Wee Keong Lim, "A simple vision-based fall detection technique for indoor video surveillance", Springer-Verlag London, 2013
7. Zafar A. Khan and Won Sohn, " Abnormal Human Activity Recognition System Based on R-Transform and Kernel Discriminant Technique for Elderly Home Care", IEEE Transactions on Consumer Electronics, November 2011, Vol. 57, No. 4
8. Stauffer, C. and W.E.L. Grimson. Adaptive Background Mixture Models for Real-Time Tracking, Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, Vol. 2 (06 August 1999), pp. 2246-252 Vol. 2.
9. Williams, A., Ganesan, D., Hanson, A. Aging in place: fall detection and localization in a distributed smart camera network. Proceedings of the 15th international conference on Multimedia, 2007, pp. 892–901
10. Vishwakarma, V., Mandal, C., Sural, S.: Automatic detection of human fall in video. Lect. Notes Comput. Sci. Patt. Recogn. Mach. Intell. 2007, 4815, 616–623
11. Chen, Y.T., Lin, Y.C., Fang, W.H.: A hybrid human fall detection scheme. Proceedings of 2010 IEEE 17th International Conference on Image Processing, 2010, pp. 3485–3488
12. Ma X, Wang H, Xue B, et al. Depth-based human fall detection via shape features and improved extreme learning machine. IEEE J Biomed Health Inform 2014; 18(6): 1915–1922. DOI: 10.1109/JBHI.2014.2304357.
13. Ni B, Wang G, and Moulin P. RGBD-HUDAACT: A color depth video database for human daily activity recognition. Consumer Depth Cameras for Computer Vision, Springer London, 2013, pp. 193–208. DOI: 10.1007/978-1-4471-4640-7_10
14. Jiang Wang, Xiaohan Nie, Yin Xia, Ying Wu and Song-Chun Zhu, Cross-view action modeling, learning and recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'14), 2014, pp. 2649–2656.
15. Hanbo Wu, Xin Ma, Zhimeng Zhang, Haibo Wang and Yibin Li, Collecting public RGB-D datasets for human daily activity recognition, International Journal of Advanced Robotic Systems July-August 2017: 1–12

AUTHORS PROFILE

L. Aneesh Euprazia received the B.E. degree in Computer Science and Engineering from St.Xavier's Catholic College of Engineering in 2009, received the M.E. degree in Software Engineering from Anna University, Trichy in 2011 . She is currently pursuing the Ph.D. degree in Information and Communication Engineering in RMD Engineering College under the supervision of Prof. K.K.Thyagharajan. Her research interests include video-based action recognition, computer vision, object segmentation, machine learning and deep convolutional neural network.

K. K. Thyagharajan received his B.Eng. degree in Electrical and Electronics Engineering from PSG College of Technology, Madras University, India and received his M.Eng. degree in Applied Electronics from Coimbatore Institute of Technology, India in 1988. He also possesses a Post Graduate Diploma in Computer Applications from Bharathiar University, India. He obtained his Ph.D. degree in Information and Communication Engineering from College of Engineering Guindy, Anna University, India in 2007. He is in teaching profession for more than three decades and served at various levels including Principal, Dean and Professor at various Engineering Colleges in Tamil Nadu-INDIA. During his tenure as a Principal and Dean he was a member of Governing Council of RMK Group of Institutions. He has written 5 books in Computing including "Flash MX 2004" published by McGraw Hill (India), which has served recommended as text and reference book by universities. He is a grant recipient of Tamil Nadu State Council for Science and Technology. He has been invited as chairperson and delivered special lectures in many National and International conferences and workshops. He is reviewer and editorial board member for many International Journals and Conferences. He is a recognized supervisor for Ph.D candidates and Master students at Anna University. He has published more than 100 papers in National & International Journals and Conferences. Nine candidates have completed PhD and eight more are doing PhD under his supervision. He has four patents published to his credit. His research interests include Computer Vision, Semantic Web, Image & Video Processing, Multimedia Streaming, Video Coding, Content-based Information Retrieval, Microcontrollers and e-learning. He is a Fellow of Institution of Engineers (India), life member of ISTE and CSI (INDIA), and also senior member and invited member in many professional associations. He has been recognized as a Teacher Par Excellence twice by the management of SSN institutions. He received Distinguished Faculty (Multimedia and Image Processing) Award from Venus International Foundation Chennai, Indo Global Education Excellence Award from International Benevolent Research Foundation Kolkata, and Best Administrator award from PEARL Foundation Madurai. He has been recognized by Marquis Who's Who in the World for his contribution to the technical society and his biography has been published in its 25th Anniversary Edition

E-mail: acdean@rmd.ac.in ,
kkthyagharajan@yahoo.com , kkthyagharajan@gmail.com
Scopus indexed publications:
https://www.scopus.com/authid/detail.uri?authorId=14046138000
Google scholar:
https://scholar.google.co.in/citations?user=R5jfVjoAAAAJ&hl=en
Research Gate: https://www.researchgate.net/profile/Thyagharajan_K_K
Publons (Web of Science):
https://publons.com/researcher/496086/thyagharajan-k-k