Real Life Human Movement Realization in Multimodal Group Communication Using Depth Map Information and Machine Learning

Dr. G. Ranganathan,
Head of the department,
Department of Electronics & Communication Engineering,
Gnanamani College of Technology,
Namakkal, India – 6370 18.
Email id: profranganathang@gmail.com

Abstract: The latest advancements in the evolution of depth map information's has paved way for interesting works like object recognition sign detection and human movement detection etc. The real life human movement detection or their activity identification is very challenging and tiresome. Since the real life activities of the humans could be of much interest in almost all areas, the subject of identifying the human activities has gained significance and has become a most popular research field. Identifying the human movements /activities in the public places like airport, railways stations, hospital, home for aged become very essential due to the several benefits incurred form the human movement recognition system such as surveillance camera, monitoring devices etc. since the changes in the space and the time parameters can provide an effective way of presenting the movements, yet in the case of natural color vision, as the flatness is depicted in almost all portions of images. So the work laid out in the paper in order to identify the human movement in the real life employs the space and the time depth particulars (Spatial-Temporal depth details –STDD) and the random forest in the final stage for movement classification. The technology put forth utilize the Kinect sensors to collecting the information's in the data gathering stage. The mechanism laid out to identify the human movements is test with the MATLAB using the Berkley and the Cornell datasets. The mechanism proposed through the acquired results proves to deliver a better performance compared to the human movements captured using the normal video frames.

Keywords: Space-Time Information's, Machine Learning, Random Forest, STDD, Histogram of Gradient

1. Introduction

Human movement observation becomes very essential in the public places like airport, railway stations, bus stops, hospital, home for aged, etc., this is becoming very essential in the home nowadays as most of the parents are employed leaving their children or the elderly in the hands of the baby sitter or care takers. The other important thing that makes the human movement identifier necessary is because the real life activities of the humans could be of much interest. The continuous and the dynamic changes taking in every area of our daily lives creates a more interest for the public as well as the researchers to understand the

actions. The sequence of images are used for this purpose rather than using the still images. Few researcher in his paper proposed the depth particulars in the identifying the activities of the humans.

The depth particulars are the image frames with the extricated per pixel details of the image in the natural form. The pixel value always varies due to the lack of flatness. The change in the details of the pixels across varies planes in the section of the images is distinguished using the depth details. This could be developed by capturing images in the different angles and views and become the primary necessity for the depth details of the images.

Variety of sensors exists to take the images depth e.g. Kinect and even more. The movements of the humans are sorted out into many sub classes based on the estimation of the pose, gesture, individual and actions performed in a group in public or in private places.

The depth details delivers minimum noise information's and improves the human movement realization systems efficiency. The space and the time particulars for extricating the features develops enormous level of multidimensional data that is highly difficult to compute without the devices that have intense resources.

The space- time are primary objectives in a research as all the features in an image vary with respect to time and space. So the paper also aims to identify the human movement in the real life employing the space and the time depth particulars (Spatial-Temporal depth details –STDD) and the random forest in the final stage for movement classification. The carried out work in the paper includes the realization of all human movements that is through the individual or the multimodal group action with the assistance of the depth particulars observed from the depth camera.

The paper below is arranged with the related works about the related studies and the past works in 2. The proposed work in 3. The results and discussion in 4 and the conclusion and the future scope in 5.

2. Related works

Kim et al [1] presents the "Human activity recognition and pattern discovery." Ye et al [2] elaborates the "survey on human motion analysis from depth data." Koresh et al [3] has performed the "Computer Vision Based Traffic Sign Sensing for Smart Transport" Anguita et al [4] devised the " public domain dataset for human activity recognition using smartphones." Vinothkanna, et al [5] presented the A Secure Steganography Creation Algorithm for Multiple File Formats. Oliver et al [6] performed the "Understanding and modeling of wifi signal based human activity recognition." Manoharan, S et al [7]

proposed a "A Smart Image Processing Algorithm for Text Recognition, Information Extraction and Vocalization For The Visually Challenged"Wang, et al [8] presented the "Layered representations for human activity recognition." Chandy, et al [9] presented the "A Review On Iot Based Medical Imaging Technology For Healthcare Applications."Zeng, et al [10] proposed the "Convolutional neural networks for human activity recognition using mobile sensors."Manoharan, S et al [11] proposed the "Image Detection, Classification and Recognition for Leak Detection in Automobiles" Luft, Thomas, et al [12] put forth the "Image enhancement by unsharp masking the depth buffer." Chandy, Abraham et al [13] has presented the "RGBD Analysis for Finding the Different Stages of Maturity Of Fruits In Farming."Henry et al [14] has devised the "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments." Bindhu, V. et al [15] has proposed the "Biomedical Image Analysis Using Semantic Segmentation." Calkoen, C. J et al [16] has put forth the . "The bathymetry assessment system: efficient depth mapping in shallow seas using radar images." Kumar, T. Senthil. et al [17] has devised the "A Novel Method for HDR Video Encoding, Compression And Quality Evaluation."

3. Proposed Work

The block diagram below in figure. Shows the stages in the work proposed in the paper for realizing the human movements using the depth details and the machine-learning.
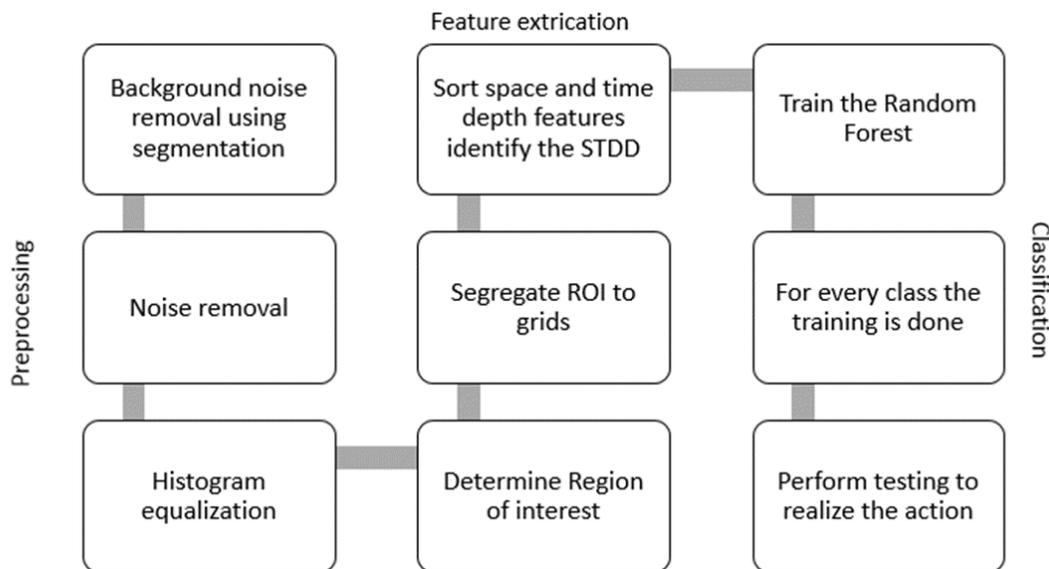

Figure.1 Block Diagram

As shown in figure.1 the work laid out is processed using three stages, first is the preprocessing, the next is the feature extrication and the third is the classification. The work laid out utilizes state of art datasets in realizing the human movements. The preprocessing is does the noise and the unwanted signal removal from the actual information this stage makes the feature extrication and the further processing easy. The second stage feature extrication is the essential and the fundamental work of the proposed process. In feature extrication every frame is processed and the regions of interest determined. This further segregated into grids of eight X five and the image is presented in the grid form. The depth details depicts every pixel $'P'$ in the image as $P = (i, j, d(i, j))'$ while $'d'$ is the depth across the $'i'$ and $'j'$ pixels. The depth is decided by the Kinect sensor used for collecting the images. The $P\ (i, j)$ is the "point on the surface. The "cross product of the tangent vector processes the normal vector for $'P'$ to the particular plane in the section of the image. The equation 1 and 2 shows the "tangent vectors" T_i , T_j

$$T_i = \frac{\partial}{\partial i} \begin{bmatrix} i \\ j \\ d(i,j) \end{bmatrix} \tag{1}$$

$$T_j = \frac{\partial}{\partial j} \begin{bmatrix} i \\ j \\ d(i,j) \end{bmatrix} \tag{2}$$

And the equation 3 gives the parameters involved in determining the "normal vectors"

$$Normal\ _{vector} = T_i\ X\ T_j = \begin{bmatrix} -\frac{\partial(i,j,d(i,j))}{\partial i} \\ -\frac{\partial(i,j,d(i,j))}{\partial j} \\ 1 \end{bmatrix} \tag{3}$$

The method engaged the dense sampling procedures to extricate the spatial and the temporal features at the usual location and the scales in the blocks of video. The information, i, j and the spatial (Sp) and the temporal (tp) are extracted from the video blocks. Integrating the dense extricated features with the optical flow based on the histogram delivers efficient outputs to assist the work laid out. The features of the human movements so far extricated are sent as input to the final stage that does the classification process. The random forest employed in the final stage classifies the features based on the "histogram of depth features (h_{df}) and dense sampling of the optical features flow (ds_{opff})" The depth $(depth_{tf})$ across the training features set are determined using the following equation 4.

$$depth_{tf} = 0.5 \, X \, \sum_{x=1}^{n} \frac{(h_{df}) + (ds_{opff})}{(h_{df}) - (ds_{opff})} \qquad (4)$$

While 'n' denotes the dimension of every samples employed in training.

4. Results and Discussion

The configuration set up for experimenting the proposed method is listed in the table.1 below. The data set used in the proposed work is gathered using the "Kinect SDK "sensor. These types of sensors are cost effective and could be installed at ease. Ten video frames based on the activities of the individual person and group with four to five person on diverse surroundings in gathered. In the information gathered every object is represented with different colors and the morphological operations is performed to elude the back ground and the noise.

Now the features from the images that are free from unwanted information's are subjected to extrication. The ROI are computed and the image is converted into eight X five images. The STDD are determined and the gap across the count of pixels that are depicting the different planes of the images are estimated. Every action is concurrent image frames are subjected to feature extraction and the h_{df} and the ds_{opff} are processed. This is fed as input to the classifier that segregates the various actions of the person the figure.2 below show the outcomes in the each stage of the proposed process.

Figure.2 Outcomes Observed

Actions	Drinking	waving hands	Talking in phone	Reading books	Typing	Stretching	Yawning
Drinking	35	0	0	0	0	0	0
Waving hands	0	33	0	0	0	0	0
Talking in phone	0	1	27	1	1	0	0
Reading book	1	3	0	31	0	0	0
Typing	0	0	3	0	28	0	0
Stretching	0	0	0	0	0	9	0
Yawning	0	0	0	0	0	5	37

Table .2 Confusion Matrix for Individual Activities

Actions	Fighting	Shaking Hands	Normal Conversation	Exchanging things	Greeting	Playing	Dancing
Fighting	650	0	0	0	0	0	0
Shaking Hands	0	330	0	0	0	0	0
Normal Conversation	0	127	275	7	27	0	0
Exchanging things	45	35	0	310	0	0	0
Greetings	0	0	23	0	285	0	0
Playing	0	0	0	0	0	495	0
Dancing	75	0	0	0	0	57	537

Table.3 Confusion Matrix for Group Activities

The table 2 and table.3 is the confusion matrix in realizing the movement of the individual activities that are tedious and easy to realize. The bitmap of the features extricated to be classified includes average depth of colors used in representing the object, the depth across the centroids which is usually one, the no.of pixels in the foreground, frame differences and region of interest grids pixel distribution. The figure .3 and 4 depicts the accuracy acquired in individual and the group actions respectively on various human movements.
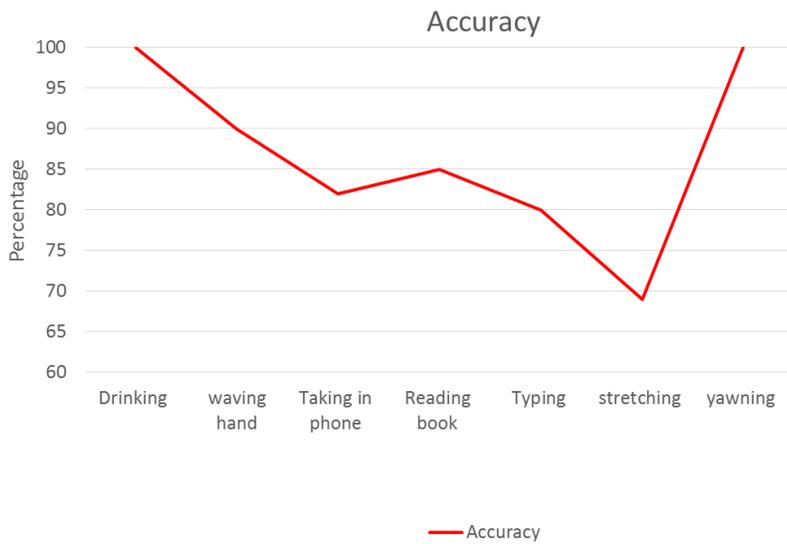
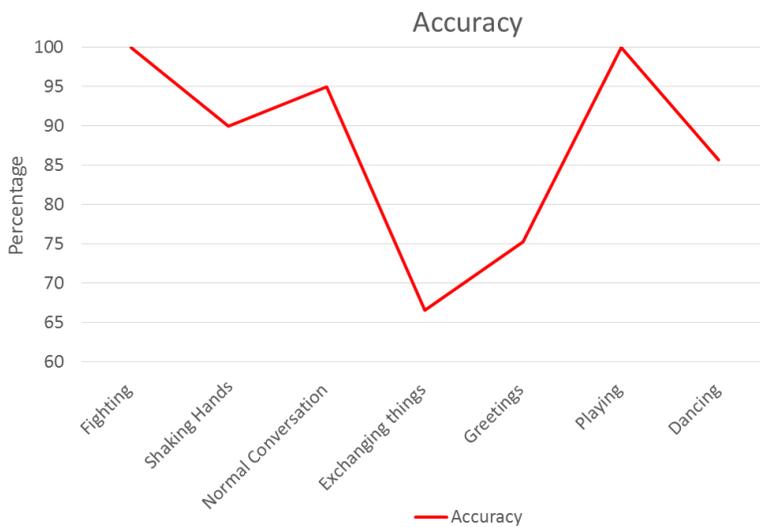Figure.3 Accuracy (Individual Actions)

Figure .4 Accuracy (Group Actions)

The Group actions are often miscalculated and improperly classified in previous methods for instance a waving hands to say good bye is often misclassified as some other actions such as throwing some object or

etc. the results observed in the confusion matrix and the accuracy acquired for proposed model shows that the depth information mapping assist better in realizing the human activities than the other models.

5. Conclusion

As the human realization in very important nowadays in multiple fields the proposed work, lay out a multimodal group human activity realization model using the, depth details and the RF classifier. The information's are collected using the Kinect sensors and preprocessed using the morphological operations, and the spatial and the temporal depth features are extricated, the bitmap of the features based on the average depth of colors used in representing the object, the depth across the centroids which is usually one, the no.of pixels in the foreground, frame differences and region of interest grids pixel distribution are provided as input to the classifier to sort out the features. The laid out frame work is not influenced by the light intensity of the surrounding environment or the luminous changes in the atmosphere. In future to subdue the difficulties in the managing the dimensionality properly the paper employs the deep learning architecture utilizing the deep network libraries and the swift optimization methods to devise an efficient realization process.

References

[1]   Kim, Eunju, Sumi Helal, and Diane Cook. "Human activity recognition and pattern discovery." *IEEE pervasive computing* 9, no. 1 (2009): 48-53.

[2]   Ye, Mao, Qing Zhang, Liang Wang, Jiejie Zhu, Ruigang Yang, and Juergen Gall. "A survey on human motion analysis from depth data." In *Time-of-flight and depth imaging. sensors, algorithms, and applications*, pp. 149-187. Springer, Berlin, Heidelberg, 2013.

[3]   Koresh, M. H. J. D. (2019). Computer Vision Based Traffic Sign Sensing For Smart Transport. Journal of Innovative Image Processing (JIIP), 1(01), 11-19

[4]   Anguita, Davide, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. "A public domain dataset for human activity recognition using smartphones." In *Esann*. 2013.

[5]   Vinothkanna, M. R. (2019). A Secure Steganography Creation Algorithm for Multiple File Formats. Journal of Innovative Image Processing (JIIP), 1(01), 20-30.

[6]   Oliver, Nuria, Eric Horvitz, and Ashutosh Garg. "Layered representations for human activity recognition." In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pp. 3-8. IEEE, 2002.

[7]   Manoharan, S. (2019). A Smart Image Processing Algorithm for Text Recognition, Information Extraction and Vocalization For The Visually Challenged. Journal of Innovative Image Processing (JIIP), 1(01), 31-38.

[8]   Wang, Wei, Alex X. Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. "Understanding and modeling of wifi signal based human activity recognition." In *Proceedings of the 21st annual international conference on mobile computing and networking*, pp. 65-76. 2015.

[9]   Chandy, A. (2019). A Review On Iot Based Medical Imaging Technology For Healthcare Applications. Journal of Innovative Image Processing (JIIP), 1(01), 51-60.

[10]   Zeng, Ming, Le T. Nguyen, Bo Yu, Ole J. Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. "Convolutional neural networks for human activity recognition using mobile sensors." In *6th International Conference on Mobile Computing, Applications and Services*, pp. 197-205. IEEE, 2014.

[11]   Manoharan, S. (2019). Image Detection, Classification and Recognition For Leak Detection In Automobiles. Journal of Innovative Image Processing (JIIP), 1(02), 61-70.

[12]   Luft, Thomas, Carsten Colditz, and Oliver Deussen. "Image enhancement by unsharp masking the depth buffer." ACM Transactions on Graphics (TOG) 25, no. 3 (2006): 1206-1213.

[13]   Chandy, Abraham. "Rgbd Analysis for Finding the Different Stages of Maturity Of Fruits In Farming." Journal of Innovative Image Processing (JIIP) 1, no. 02 (2019): 111-121.

[14]   Henry, Peter, Michael Krainin, Evan Herbst, Xiaofeng Ren, and Dieter Fox. "RGB-D mapping: Using Kinect-style depth cameras for dense 3D modeling of indoor environments." The International Journal of Robotics Research 31, no. 5 (2012): 647-663.

[15]   Bindhu, V. "Biomedical Image Analysis Using Semantic Segmentation." Journal of Innovative Image Processing (JIIP) 1, no. 02 (2019): 91-101.

[16]   Calkoen, C. J., G. H. F. M. Hesselmans, G. J. Wensink, and J. Vogelzang. "The bathymetry assessment system: efficient depth mapping in shallow seas using radar images." *International Journal of Remote Sensing* 22, no. 15 (2001): 2973-2998.

[17]   Kumar, T. Senthil. "A Novel Method for HDR Video Encoding, Compression And Quality Evaluation." Journal of Innovative Image Processing (JIIP) 1, no. 02 (2019): 71-80.

Authors Biography

Dr. G. Ranganathan, is currently the Head of the department, in Department of Electronics & Communication Engineering, at Gnanamani College of Technology, in Namakkal, India. His major areas of research are Image Analysis, Object tracking and Recognition, Image coding, enhancement, Filtering, Rendering, Restoration, Half-toning, Search, and Vision-based Human - Computer Interaction.