# A Mono Master Shrug Matching Algorithm for Examination Surveillance

**Sandhya Devi G[1], Prasad Reddy P V G D[2], Suvarna Kumar G[3], Vijay Chaitanya B[4]**
[1]Department of Computer Science & Systems Engineering, Andhra University, India
[2]Department of Computer Science & Systems Engineering, Andhra University, India
[3]Department of Computer Science & Engineering, MVGR College of Engineering, India
[4] Department of Computer Science & Engineering, MVGR College of Engineering, India

*Abstract*—This paper proposes an unusual slant for Shrug recognition from Gesticulation Penetrated Images (GPI) based on template matching. Shrugs can be characterized with image templates which are used to compare and match shrugs. The proposed technique makes use of a single template to identify match in the candidates and hence entitled as mono master shrug matching. It does not necessitate erstwhile acquaintance of movements, motion estimation or tracking. The proposed technique brands a unique slant to isolate various shrugs from a given video. Additionally, this method is based on the reckoning of feature invariance to photometric and geometric variations from a given video for the rendering of the shrugs in a lexicon. This descriptor extraction method includes the standard deviation of the gesticulation penetrated images of a shrug. The comparison is based on individual and rational actions with exact definitions varying widely uses histogram based tracker which computes the deviation of the candidate shrugs from the template shrug. Far-reaching investigation is done on a very intricate and diversified dataset to establish the efficacy of retaining the anticipated method.

*Index Terms*— Gesture Recognition, Template Matching, Video Surveillance, Suspicious Activity Detection

## I. INTRODUCTION

Shrug is basically a combined moment of the shoulder and the elbow or more precisely a shoulder gesture which is the raise of one's shoulders slightly and momentarily that gesticulates indifference. The shrug matching system can be instigated in an eclectic assortment of applications ranging from traffic monitoring for the development of intelligent surveillance systems. Now a days automatic shrug matching is widely applied in the fields of bio-mechanics, sports, cinema, etc. for cutting-edge scrutiny.

Humanoid gesticulations are decisive ever since the machines are obligatory to intermingle more and more perceptively and painlessly with a humanoid tenanted milieu. In order to mend the real-time machine proficiencies, it is imperative to epitomize gesticulation. Conversely owing to innumerable precincts, no single approach appears to work sufficiently in understanding and recognizing actions. Template matching methodologies entail meeker and sooner algorithms for gesticulation scrutiny that can epitomize an intact video sequence into a single image format.

In this paper, a competent Mono-master shrug matching approach was urbanized based on a sequence of Gesticulation Penetrated Images (GPI). This encompasses the use of GPI obtained utilizing infrared depth sensors. Motion based template matching techniques are labored on training videos to attain the crucial topographies from the gesticulation sequences. The feature vector is fashioned by commissioning statistical maneuvers and spatial-temporal motion information. The testing phase is alienated into several paces. Initially diverse gesticulations are alienated from the protracted test sequences. Then, analogous to the train feature vectors test feature vectors are engendered for each gesticulation. Finally, every test feature vector is paralleled to each train feature vector for dissimilar gesticulations and a classifier is labored to find the best possible match of a gesticulation from the given training lexis.

## II. RELATED WORKS

Gesticulations are a business of far-reaching investigation and a number of techniques have been industrialized (Ahad, 2011), which are by now being instigated in selected concrete solicitations. Slants differ predominantly on the basis of depiction of action, which is universally performed based on advent, shape, spatial-temporal orientation, optical flow, interest-point and volume. In recent times, various researchers used optical flow-based motion detection and localization methods and frequency domain depiction of gesticulations. In spatial-temporal template based methodologies an image sequence is used to formulate a Motion Energy Image (MEI) and a Motion History Image (MHI) which signpost the regions of motions and realm the time information of the motion as well (Ahad et al., 2008; Bobick and Davis, 2011).

The model-based methodology on the other hand ascertains the body limbs and tracks those using a 2D model of the body. The models are significantly based on pose estimations. Well-known models include patches, cylinders, and clay patterns (Ahad, 2011). Humanoid gesticulations are fragmented into atomic parts, demonstrating elementary movements.

The approaches fluctuating from effortless techniques such as frame differencing and adaptive median filtering, to more sophisticated probabilistic modeling techniques (Cheung and Kamath 2004) are used for the reason that

convoluted techniques habitually fabricate superior recital, experiments show that effortless techniques such as adaptive median filtering can produce fine results with much lower computational complexity. Back ground extraction from video sequences (Srenivas Varadarajan, Lina J. Karam) the still back ground occluded by number of foreground objects is done. This algorithm is proficient and effortless and be able to yet recuperate the background from the anticipated resource frames as a substitute performing every time the fore ground background taxonomy.

In diverse data sets, soaring precisions have been attained by numerous number of existing gesture recognition methods. If the training data are restricted they do not carry out sound, nevertheless the majority of them do depend on fine measure of input to train the system.

Some algorithms have been building up to seize benefit of the profundity information with an escalated depth sensor from the liberation of the Kinect sensor. Several works accomplished something in making out the body elements and exploiting a sensor in tracking them to make out actions and gestures.

For the reason that small set of gestures was used with a very little variation in recognition of gesticulations was simpler. These gestures probably include the extending of hands in front of the body to ease up the detection.

### III. Proposed method for Mono-Master Shrug Matching

In this paper, a method of shrug matching using a single template from a trifling lexis of gesticulations is anticipated. Every application prerequisites a dedicated gesticulation lexis. In recent years gesture recognition has turned out to be a part of our diurnal and henceforth entail gesture recognition engines which can certainly get bespoke to new gesture lexes. The gesticulations will be pinched from a trifling Lexis of gesticulations largely related to a particular task, for instance, the moment of limbs and hands with a student in an examination hall.

Equipping a slice of training specimens may be impractical in voluminous applications where recording data and labeling them is monotonous. Examination surveillance using shrug matching will be conceivable only if systems can be trained to recognize new shrugs with very few and perfectly one. In our work, it is assumed that in a given data set consist of both RGB and GPIs. The proposed algorithm focuses only on the depth data from the sensor for the shrug matching model.

### IV. Shrug Dataset

There are some clearly defined hand, body or head gesture datasets like NATOPS aircraft handling signal database, Cambridge gesture dataset. All of these existing datasets are used to identify the type of communication such as signals, signs on sign boards and so on. Nonetheless eminent for their insides and intricacies, all

of these datasets report a precise brand of gesticulations restricted to a number of classes and solicitation dominions. Consequently, for the mockup tenacity of the anticipated method, a rich but exceedingly intricate dataset was primed and deliberated exclusively for examination surveillance. The gesticulation database encompasses three groupings of gesticulations analogous to different poses of a candidate appearing for an examination. The categories include (1) Shrugs in the customary pose of writing the examination, (2) Stretching of arms towards the other worktable, (3) Leaning down on the worktable. The dataset comprises of videos from RGB and depth cameras of a Microsoft Kinect Sensor (Fig. 1). Each set of data encompasses a number of shrugs offered discretely and just once for training tenacity.

In the videos, an examination hall is portrayed in front of a fixed camera. The videos are a collection of a dataset of an examination hall, including RGB and depth videos recorded with Kinect camera with image sizes of 640 x 480 pixels at 10 fps (frames per second). The environment is set to be different and some contain poor lighting. The stances of the body are unlike in different sets, habitually exposing the upper frame of the body, but some even spectacle only a part of the body or it is shown from behind. In addition, in numerous circumstances the videos are imperiled to self-occlusion.



Fig. 1. Shrug Dataset: RGB and corresponding GPIs'

### V. Feature Extraction and Training

The depth data provided in the dataset are in RGB format and therefore need to be converted to grayscale before processing. The grayscale depth data is an exact depiction of the object distance from the camera by variable intensity of pels from dark to bright for near and far objects, correspondingly. The binary image is cast-off as a mask to clean out the background from the GPI and by this means separate the human object from the image (Fig. 2).

The proposed method of shrug matching extremely focuses on three types of operations for obtaining the match measure between the master and the candidate images from the training videos. The operations are (a) performing dilation and erosion, (b) performing standard

deviation on each pixel value across time; (c) histogram based comparison on each pixel value.

Dilation and Erosion are the rudimentary maneuvers in morphological image processing which were formerly urbanized for binary images and later stretched to grayscale and to complete lattices. The dilation and erosion operations customarily use a structuring element for probing and intensifying the shapes confined in the input image.

A binary image is a subset of Euclidian space $R^d$ for a dimension $d$. Let S be a subtracting element for an image I, then the dilation of an image I by the structuring element S is given by

$$I \oplus S = \bigcup_{s \in S} I_s \tag{1}$$

Consider the input matrix from a binary image before dilation

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{2}$$

After performing dilation the matrix of the binary image is as follows

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{3}$$

And the erosion of an image I by the structuring element S is given by

$$I \ominus S = \bigcap_{s \in S} I_s \tag{4}$$

Consider the input matrix from a binary image before erosion

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{5}$$

After performing erosion the matrix of the binary image is as follows

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \tag{6}$$



Fig. 2. GPI and its corresponding Masked and Background subtracted image

Standard deviation is a very modest but commonly used statistical degree. It mines the information on deviation of values from the mean in an array. In shrug matching, the values of the STD performed on each pel for certain time duration can be utilized to track the change in the candidate's position within that time. It engenders an image quite analogous to the motion energy image (MEI) but with color tones depicting hasty to relaxed vagaries by varying from red to blue regions. For the $n^{th}$ gesture at lexicon G consisting of F frames, the standard deviation ${}_F\sigma_n^2(x, y)$ of pel (x, y) across the frames is given by,

$$_F\sigma_n^2(x, y) = \frac{\sum (I_{xy}(t) - \hat{I}_{xy})^2}{F} \tag{7}$$

Here $I_{xy}(t)$ is the pel value of the location (x, y) of the frame at time t, where x=0,1, 2, 3, . . . , p, y=0, 1, 2, 3, . , q and t=0, 1, 2, 3, . . . , F. $\hat{I}_{xy}$ is the mediocre of all $I_{xy}(t)$ values along time t. Therefore, for the whole frame, the matrix obtained is defined as,

$$_F\Delta_n^G = \begin{pmatrix} _F\sigma_n^2(0, 1) & \cdots & _F\sigma_n^2(0, q) \\ \vdots & \ddots & \vdots \\ _F\sigma_n^2(p, 1) & \cdots & _F\sigma_n^2(p, q) \end{pmatrix} \tag{8}$$

The matrix itself can be cast-off as a feature for each of the lexicon G. The matrices are acquired by carrying out standard deviation on the training samples of development data of the shrug dataset is shown in fig. It is apparent from fig that carrying out standard deviation (STD) across frames boosts the evidence of drive across the frames while subduing the static chunks. One can effortlessly gain an intuition about track of motion flow and brand of the shrug by discerning these figures. Thus, STD across frames is one of the anticipated feature vector development methods.

## VI. FINDING SHRUG BOUNDARY

The dataset provides the test execution samples containing a set of shrugs. Therefore, the first task is to separate the different shrugs from the shrug sequences. Thereafter, the task involves the subtraction of the each candidate frame from the master frame is used to find the shrug boundary.
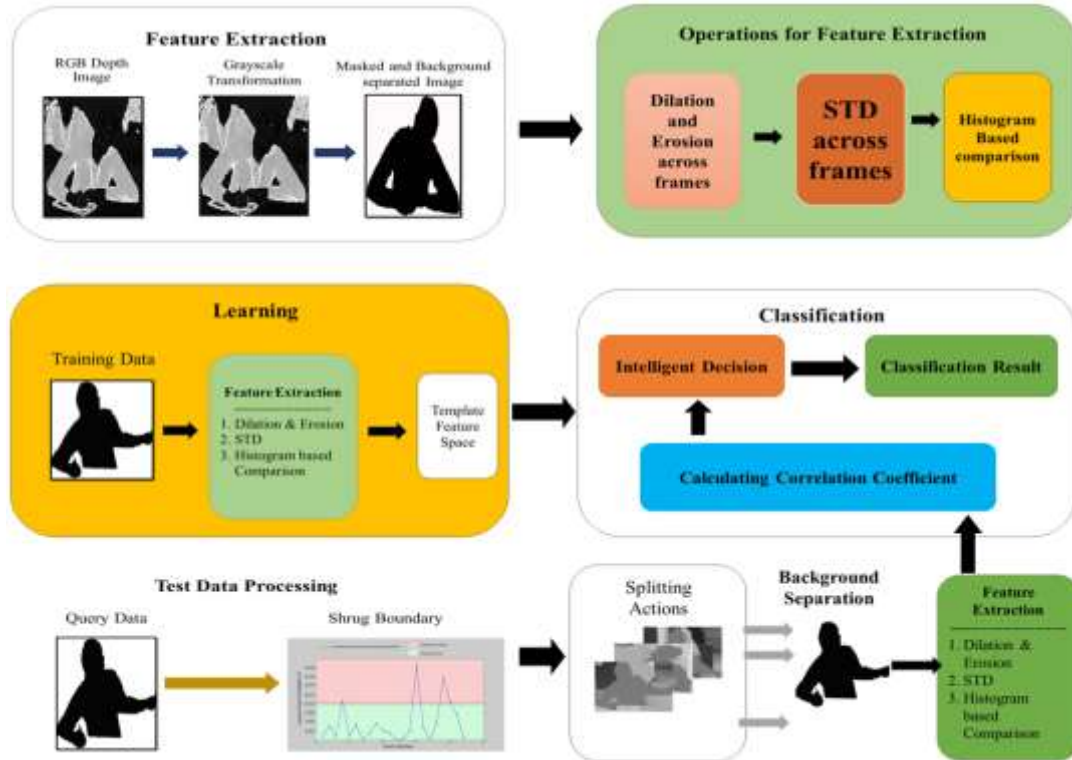
Fig. 3. Proposed method for Shrug Matching

For a p x q gesticulation penetrated frame, if $I_{xy}(t)$ is the pel value of the position (x, y) at a time interval t. Here t = 1, 2, 3, . . , F, then the formula mentioned below can be used to generate a variable $\alpha_t$, which provides an amplitude of change from the Master frame.

$$\alpha_0 = (I_{00}(t) - I_{00}(f0))^2 + (I_{01}(t) - I_{01}(f0))^2 + . . + (I_{0q}(t) - I_{0q}(f0))^2 \qquad (10)$$

$$\alpha_1 = (I_{10}(t) - I_{10}(f0))^2 + (I_{11}(t) - I_{11}(f0))^2 + . . + (I_{1q}(t) - I_{0q}(f0))^2 \qquad (11)$$

$$\alpha_2 = (I_{20}(t) - I_{20}(f0))^2 + (I_{21}(t) - I_{21}(f0))^2 + . . + (I_{2q}(t) - I_{2q}(f0))^2 \qquad (12)$$

…………

…………

$$\alpha_p = (I_{p0}(t) - I_{p0}(f0))^2 + (I_{p1}(t) - I_{p1}(f0))^2 + . . + (I_{pq}(t) - I_{pq}(f0))^2 \qquad (13)$$

Thus, the sum of the square differences between the candidate frame and the master frame can be obtained by,

$$\alpha_t = \sum_{n=0}^{p} \alpha_n \qquad (14)$$

From the (Fig. 4) it is depicted that each ridge in these curves represents a shrug value. Using the locations of these ridges in time along with some brainy policymaking techniques on the curve, the shrug boundaries can be found with best exactitude.
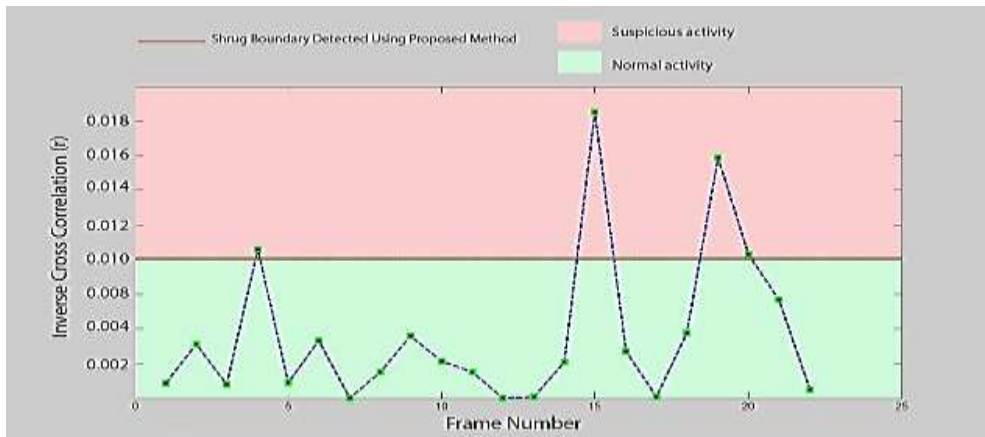


Fig. 4 Shrug Boundary detection for the dataset

## VII. CLASSIFICATION

Subsequent to dig out the features from the shrugs in the training dataset of a particular lexicon, a feature vector table is molded for that lexicon as shown in the fig. Then for the test mockups, first the gestures are unglued if multiple gesticulations exist. Then features are pulled out for each gesticulation in a modus analogous to that done in the training mockups.

For gesticulation series of frame size p x q, the feature vector gained from the test gesticulations are also p x q matrices for each brand of features. Later the difference between the master shrug to that of the candidate gestures is obtained based on Histogram based tracker by calculating the Inverse correlation coefficient between the master histogram and the candidate histogram. The Inverse correlation coefficient is a mathematical measure of the contrary relationship between two variables such that they move in opposite directions. An inverse correlation is denoted by the correlation coefficient r having a value between -1 and 0, with r = -1 indicating perfect inverse correlation. It is an extensively used measure for gesticulation recognition and template matching (Gonzalez and Woods, 2001). The inverse correlation coefficient between the master M and the candidate C is given by,

$$r = \frac{\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}(M_{xy}-\bar{M})(C_{xy}-\bar{C})}{\sqrt{\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}(M_{xy}-\bar{M})^2}\sqrt{\sum_{x=0}^{p-1}\sum_{y=0}^{q-1}(C_{xy}-\bar{C})^2}} \quad (15)$$

Where, $M_{xy}$ is the intensity of pixel (x, y) in master M and $C_{xy}$ is the intensity of the pixel (x, y) in candidate. $\bar{M}$ and $\bar{C}$ are the respective means of the intensities of all the pixels in the master and candidate. If r = -1 then there is a strong negative (inverse) correlation. A low value of the inverse correlation between the test samples to that of training samples indicates the shrugs to be identical. Further a decision is taken to ascertain the preeminent match.

## VIII. EXPERIMENTAL RESULTS

To estimate shrug matching and its performance, an extensive dataset was prepared suiting to the application of examination surveillance. The proposed takes its form in the procedure cited as below

1. Captivating the Gesticulation Penetrated Images (GPI) and performing a gray scale threshold on the same.
2. Masking the so obtained GPI and applying the background subtraction based on the depth values, in order to captivate the humanoid objects.
3. Performing the below mentioned operations for feature extraction.
   i) Applying the morphological operations namely dilation and erosion in order to intensify the features.
   ii) Performing standard deviation (STD) across the master and candidate frames at pixel level to estimate the acceleration of the candidate across several frames.

4. Finding the shrug boundary from the summation of the values of standard deviation.
5. Calculating the Inverse cross correlation between the histograms of the STD's of the master and the candidates.

Table 1. Finding Shrug Boundary and making Intelligent Decision

| Template | Candidate | r Value | Decision |
|---|---|---|---|
| s-bw-p1 | s-bw-p2 | 0.0044 | Normal |
| s-bw-p1 | s-bw-p3 | 0.00157 | Normal |
| s-bw-p1 | s-bw-p4 | 0.0040 | Normal |
| s-bw-p1 | s-bw-p5 | 0.0430 | Normal |
| s-bw-p1 | s-bw-p6 | 0.0046 | Normal |
| s-bw-p1 | s-bw-p7 | 0.00166 | Normal |
| s-bw-p1 | s-bw-p8 | 4.9467e-05 | Normal |
| s-bw-p1 | s-bw-p9 | 0.0076 | Normal |
| s-bw-p1 | s-bw-p10 | 0.0106 | Normal |
| s-bw-p1 | s-bw-p11 | 0.0076 | Normal |
| s-bw-p11 | s-bw-p12 | 4.4820e-05 | Normal |
| s-bw-p1 | s-bw-p1 | 0.0105 | Suspicious |
| s-bw-p1 | s-bw-p1-m | 0.0135 | Suspicious |
| s-bw-p1 | s-bw-p2-m | 0.0189 | Suspicious |
| s-bw-p1 | s-bw-p3-m | 0.0694 | Suspicious |
| s-bw-p1 | s-bw-p4-m | 0.0416 | Suspicious |
| s-bw-p1 | s-bw-p5-m | 0.0285 | Suspicious |

In **TABLE 1** the proposed results of mono master shrug matching algorithm for examination surveillance, where in the intelligent decision is justified based on the value of **r (inverse cross correlation)** are summarized. Our data set consists of a mono master (s-bw-p1) and the corresponding candidates are experimented for scrutiny and the results are plotted for finding the shrug boundary as shown in the Fig. 4.

## IX. CONCLUSION

In this paper, an unusual slant for shrug matching is proposed. It employs combinations of statistical measures, morphological operations and depth data. The proposed method first utilizes parameters extracted from gesticulation penetrated images (GPI) followed by the subtraction of the background. Then, features from each shrug sequence are extracted based on the operations of standard deviation across frames, performing morphological operations such as dilation followed by erosion for shrug matching. The usefulness of applying inverse correlation coefficient is apparent from the results.

### REFERENCES

[1] Ahad, M.A.R., 2011. Computer vision and action recognition: a guide for image processing and computer vision community for action understanding.

[2] Atlantis Ambient and Pervasive Intelligence. Atlantis Press. Ahad, M.A.R., Tan, J., Kim, H., Ishikawa, S., 2008. Human activity recognition: various paradigms. In: Internat. Conf. on Control, Automation and Systems. ICCAS 2008, pp. 1896–1901. http://dx.doi.org/10.1109/ICCAS.2008.4694407.

[3] Gonzalez, R.C., Woods, R.E., 2001. Digital Image Processing, Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[4] Upal Mahbub, Hafiz Imtiaz, Tonmoy Roy, Md. Shafiur Rahman, Md. Atiqur Rahman Ahad, 2013. A Template matching approach of one-shot-learning gesture recognition.

[5] Bobick, A.F., Davis, J.W., 2001. The recognition of human movement using temporal templates. IEEE Trans. Pattern Anal. Machine Intell. 23 (3), 257–267. http://dx.doi.org/10.1109/34.910878.

[6] Imtiaz, H., Mahbub, U., Ahad, M., 2011. Action recognition algorithm based on optical flow and RANSAC in frequency domain. In: Proc. SICE Annual Conf. (SICE), pp. 1627–1631.

[7] Min Li, Zhaoxiang Zhang, Kaiqi Huang and Tieniu Tan, "Estimating the Number of People in Crowded Scenes by MID Based Foreground Segmentation and Head-shoulder Detection", IEEE Computer Society PressVol. 35, pp.96-120, 2008.

[8] H. Zhou, Y. Yuan and C. Shi, "Object Tracking Using SIFT Features and Mean Shift," *Computer Vision and Image Understanding*, Vol. 113, No. 3, 2009, pp. 345-352.

[9] Handbook of Image and Video Processing, By Alan Conrad Bovik, Al Bovik Contributor AlBovik Edition: 2, illustrated, revised Published by Academic Press, 2005.

[10] B V Ramana, "Engineering Mathematics", TATA McGraw Hill, 2002.

[11] Gonzalez, R, Woods, R, Eddins, S "Digital Image Processing using Matlab" Prentice Hall, 2004.

[12] Yuan, Po, M.S.E.E. "Translation, scale, rotation and threshold invariant pattern recognition system". The University of Texas at Dallas, 1993, 62 pages; AAT EP13780.

[13] H. Y. Kim and S. A. Araújo, "Grayscale Template-Matching Invariant to Rotation, Scale, Translation, Brightness and Contrast," IEEE Pacific-Rim Symposium on Image and Video Technology, Lecture Notes in Computer Science, vol. 4872, pp. 100-113, 2007.

[14] Shao, L., Chen, X., 2010. Histogram of body poses and spectral regression discriminant analysis for human action categorization. In: Proc. British Machine Vision Conference. BMVA Press, pp. 88.1–88.11. http://dx.doi.org/10.5244/C.24.88.

[15] Shao, L., Wu, D., Chen, X., 2011. Action recognition using correlogram of body poses and spectral regression. In: 18th IEEE Internet Conf. on Image Processing (ICIP), pp. 209–212. http://dx.doi.org/10.1109/ICIP.2011.6116023.

[16] Shao, L., Ji, L., Liu, Y., Zhang, J., 2012. Human action segmentation and recognition via motion and shape analysis. Pattern Recognition Lett. 33 (4), 438–445. http://dx.doi.org/10.1016/j.patrec.2011.05.015.

[17] Willamowski, J., Arregui, D., Csurka, G., Dance, C., Fan, L., 2004. Categorizing nine visual classes using local appearance descriptors. In: Workshop on Learning for Adaptable Visual Systems. In: IEEE International Conf. on Pattern Recognition.

[18] Wu, D., Shao, L., 2012. Silhouette analysis based action recognition via exploiting human poses. In: IEEE Trans. Circuits and Systems for Video Technology.

[19] Yang, W., Wang, Y., Mori, G., 2009. Human action recognition from a single clip per action. Learning, 482–489.

[20] Sen-Ching S. Cheung and Chandrika Kamath [2004] Robust techniques for background subtraction in urban traffic video,fcheung11, kamath2g@llnl.gov,Center for Applied Scientific Computing, Lawrence Livermore National Laboratory

[21] Srenivas Varadarajan*, Lina J. Karam*, and Dinei Florencio, BACKGROUND RECOVERY FROM VIDEO SEQUENCES USING MOTION PARAMETERS ,Department of Electrical Engineering, Arizona State University, Tempe, AZ 85287-5706 , Microsoft Research, One Microsoft Way, Redmond, WA 9805

**Authors' Profiles**

**Sandhya Devi G** is currently a part time research scholar in CS&SE department of Andhra University, India. She has 2 years of teaching experience. Her current research interests include image processing and analysis, design of surveillance systems for various applications, machine learning. She published 3 international journals.

**Dr.Prasad Reddy P.V.G.D** is currently the Rector (Pro-vice chancellor) of Andhra University. He completed his Ph.D. in the year 1993. He has 24 years of teaching and research experience which includes more than 10 years of administrative experience with Andhra University at various senior capacities.

He has so far successfully guided 12 Ph.D.'s, 18 M.Phil.'s and 259 M.Tech Projects in the field of Computer Engineering, and has to his credit, more than 104 research papers, which include 42 publications in international journals of repute, 2 patents filed and he has presented 19 publications in international conferences across the world.

**Dr.Suvarna Kumar G** is currently working as the Associate Professor in CSE Department of MVGR College of Engineering. His Research area is Speech and video processing.

He published 14 international journals and attended 4 conferences in Singapore. He has 12 years teaching experience, 2 years R &D and 1 year industrial experience. He is a member of IEEE, CSI,ACM, National Science Congress, IAENG and IACSIT.

**Vijay Chaitanya B** received the B.Tech degree in Computer Science & Engineering from MVGR College of Engineering, Andhra Pradesh, India, in 2014. His current research interests include computer vision, pattern recognition, digital image processing and analysis, multimedia, and visual surveillance. Currently he is a Software Developer at HCL Technologies, India.

Mr. Vijay Chaitanya B is the member of the Computer Society of IEEE, Oracle Certified Professional and Microsoft Certified Professional.