

SURVEILLANCE VIDEO BASED ROBUST DETECTION AND NOTIFICATION OF REAL TIME SUSPICIOUS ACTIVITIES IN INDOOR SCENARIOS

Nithya Shree R, Rajeshwari Sah and Shreyank N Gowda

Department of Computer Engineering,
R. V. College of Engineering, Bangalore, India
nithyashree675@gmail.com
rajeshwari.sah@gmsil.com
kini5gowda@gmail.com

ABSTRACT

Over recent years, surveillance camera is attracting attention due to its wide range of applications in suspicious activity detection. Current surveillance system focuses on analysing past incidents. This paper proposes an intelligent system for real-time monitoring with added functionality of anticipating the outcome through various Image processing techniques. As this is a sensitive matter, human decisions are given priority, still facilitating limited logical intervention of human resource. This framework detects risk in the area under surveillance. One such dangerous circumstance is implemented, like a person with a knife. Here the prediction is that in the firm places like ATM, Banks, Offices etc. a person possessing knife is unusual and likely to cause harmful activities like threatening, injuring and stabbing. The experiment demonstrates the effectiveness of the technique on training dataset collected from distinct environments. An interface is developed to notify concerned authority that boosts reliability and overall accuracy.

KEYWORDS

Surveillance behaviour; Real-time dynamic video; pattern match; precautionary measures; sharp objects;

1. INTRODUCTION

The rapid growth of criminal cases have increased the need to establish image processing technology in security based system. Surveillance camera is an integral part of any threat monitoring system in various public scenarios like conferences, shopping mall, restaurants, community gathering etc. Hence, surveillance input plays an intuitive role in abnormal activity detection. Moreover, digital equipment such as Web camera, processing machine instance and hard disk drive are mass-produced, and are sold at low price [2]. For analysis of real time events, multiple frameworks are proposed; they include tracking, learning and monitoring surveillance footage [1-4]. Such analysis concentrates on factors like human posture [5-8] [13], hand Jan Zizka et al. (Eds) : CCSEIT, AIAP, DMDDB, MoWiN, CoSIT, CRIS, SIGL, ICBB, CNSA-2016 pp. 227–236, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.60618

movements [9-12] and object properties present in the video frame sequences [8]. To estimate the possibility of danger caused by any physical force is well identified in this paper by recognizing the sharp and harmful weapons, hand prehensile and grasping movements. A person being attacked by a knife in an environment of more than one individual is the major concern of our research project.

The proposed framework is a novel approach to detect and notify indoor violence like stabbing, thrashing or any activity involving physical force, to the concerned authority. The implementation is useful for examining such suspicious activity in enclosed places like ATM, Classroom, theatre, houses etc. A drastic change is brought in the society by introducing technology in enhancing the security system prevailing in the present world. The system is also dynamic in sending message to the relevant authority for undertaking precautionary measures or enforcing the desired action.

2. RELATED WORK

The influencing factor to the experiment is the benefit of social security that is foreseen by enhancing the functionality of Visual surveillance camera by embedding the provision of processing, detecting and notifying the menace befallen by the person who is captured in its input video.

We are interested in the previous work that come from the field of human interaction and motion tracking in order to qualitatively analyse the real time human motion and predict the ongoing activities in contact with other human being present and with the objects. Most of these work use background subtraction practice to get the binary foreground image. [5][6][3].

A vast amount of literature exists on real time hand tracking. Few are picked as follows: Javeria Farooq et al. [12], Milad Rafiee Vahid et al. [16], P Raghu Veera Chowdary et al. [9], Mykyta Kovalenko et al. [10] and Pedro Cisneros et al. [11] provide information to the logic of the experiment for capturing the real time hand movement data and processing them using matching, computing and anticipation techniques to get the actual hand behavioural understanding that helps in decision making of the scenario under test. We are interested in works from pattern recognition [13] and comparison for identifying the hand posture and comparing it with the data base images [14]. Also, the knowledge with respect to hand prehensile and clutching movements are acquired that is used in the apprehensive area for identifying the hand mould to conclude the presence of object in hand.

Although, Human activity processing and hand-object interaction analysis is a difficult task from a machine vision perspective it has extensive benefit to the Societal security and brings a drastic awareness in the current system to prevent offensive conduct in public space.

The approach is to recognise the presence of sharp object in hand which is one of the means to cause harm. The shape, orientation and projection of the hand held object is to be discovered. Radu-Daniel Vatavu et al. [15] conducted an investigation on the feasibility of using the posture of the hand during prehension in order to identify geometric properties of grasped objects such as size and shape. Garg et al. [8] suggested an image segmentation technique to recognize the hand object in a scene using Mathematical analysis and locate the object position in the Scene. However most of the work have concentrated on hand holding objects in general but not

specifically sharp and harmful objects like knife. We intend to design a model which identifies such cases and also notify in case of any threat.

3. METHODOLOGY

The block diagram of proposed suspicious activity detection framework is shown in Fig 1. The proposed framework consists of two main components: 1. (FBD) Framing and Blob detection (Input video processing) 2. (HON) Human tracking, Object (like knife) identification and notification (query Image processing) stage.

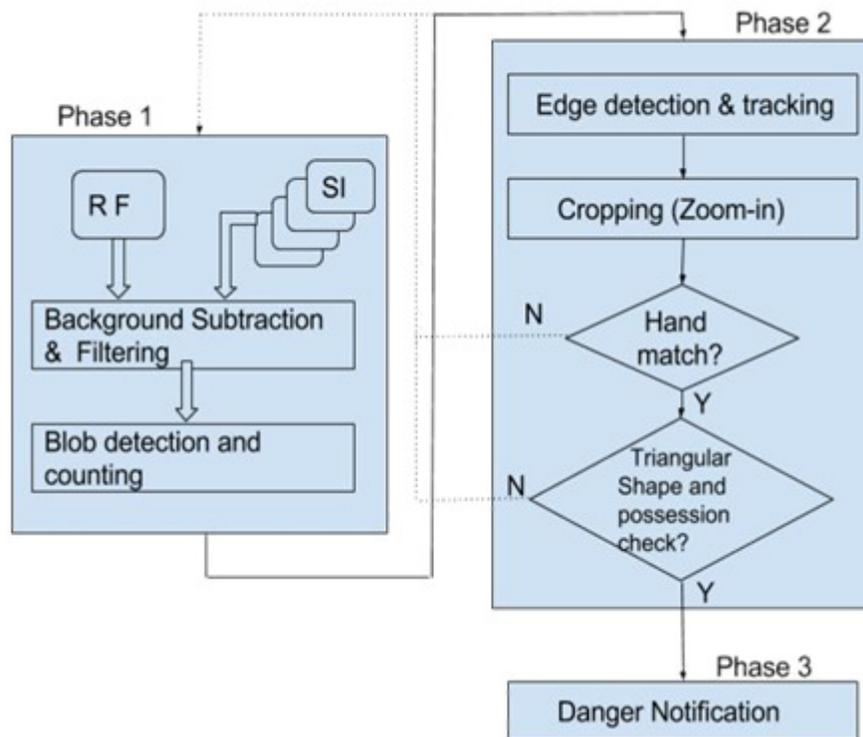


Figure 1. Block Diagram of complete algorithm

In the Framing and Blob detection stage as shown in Phase 1 of block diagram in Fig. x, Frames are extracted from the input surveillance video and compared with respect to the reference background image. Also the blob detection and analysis is done to calculate the number of blobs (i.e., human being in concern). If the number of blobs are greater than one then the program enters the next section i.e., Phase 2 of the block diagram shown in fig. x, otherwise the execution flow comes to an end for the particular frame and continued with next consecutive frame in the video.

In the Human tracking, Object (knife) identification and notification stage, edge detection helps in observing the characteristics of human blob present in the scene. The frame is zoomed-in concentrating the blobs by cropping the frame into certain dimension and is further divided into 'x' regions (and stored in the current working directory). Every block is further processed for accuracy.

Object detection returns positive result if the three conditions are satisfied; 1. Hand mould and gesture verification 2. Presence of Sharp object affirmation and 3. Link between the hand and object to meet the desired soleness. If the scenario is validated to be insecure, the program flow reaches the Phase 3 of the block diagram shown in fig. x in which a notification along with relevant photograph or latest timestamped part of video is sent to the concerned security system like nearby Police department, guard, NGO or anyone who can take quick action.

3.1 Framing and Blob detection (FBD)

The input video is taken from the surveillance camera with the Camera specification being 2MP. The video is sliced to form frames and is stored in a directory 'frames' within present working directory. A precise initial frame n1 is taken as a base frame with reference to which the further analysis of the video is done, it is basically the background image.

The looping construct is used in order to traverse the complete set of frames of the video with N_i frames where i from 2 to number of frames (n) and each frame is inspected against the base frame. Each frame undergoes the blob detection step. Blob detection is achieved through typical background subtraction between the current frame and background base frame. Morphological operations are performed in order to remove the holes, noise and unneeded information from the subtracted image indicates the number of individuals in the frames which can be determined by the area calculation method where in a defined threshold is set for the white fields appearing on the image to become blobs. If a particular white patch meets the threshold set then that particular object is considered as blob. If the number of blobs in a frame exceeds 2, then the further processing is continued. Otherwise the process halts anticipating that no hazard can take place as there is only one person, hence returns to continue with the next frame in the sequence of the video consecutively. The algorithm is efficiently implemented in order to optimize the resource usage.

3.2 Human tracking, Object (knife) identification and notification (HON)

Edge detection using Sobel's algorithm helps in human being tracking and systematic user interface design. The blobs are cropped to ignore the unnecessary processing of complete frame. The cropping is done using a set of if-else-if ladder along with the decision making factors being brute force length-width ratio calculation having known the centroid of the blob, originate pixel and area. The cropped blob is stored in the current working directory of the running program.

The blob concentrated image is further divided into multiple regions to achieve high accuracy and is accessed from the same PWD (present working directory). The cropped images are further taken for analysis.

The next step takes the image parts for processing and checks to see if these 3 conditions are contented. Only if the 3 below noted criteria's are met, the system tends to notify the desired end user and further action is left to the security system in charge. The conditions are as follows:

1. Hand mould and gesture verification: The hand posture is evaluated by matching with the data base images in the repository in complete 360 degrees to see if it is of clenched fist form. This phase concludes that the hand is holding some object.

2. Triangular or sharp object detection: The block of cropped image is checked to identify sharp object present in the frame. This is achieved by performing triangular object detection test on the frame by applying certain range of threshold condition on the length-width ratio, type of triangle, orientation of triangle etc. This ensures the presence of sharp object (knife like object) present in the frame.
3. There should be a link between hand suggested by condition 1 and possessed object suggested in condition 2. This framework is not concerned about the presence of sharp object alone, it should be held by any of the person in the frame. This can be verified by the logic with soleness check of the object in the image, i.e., there should be a single speck in the binary frame to confirm that the object x is possessed by the person only.

If all the conditions are satisfied and the flow of execution reaches till the end overcoming every decision making construct, then it can be concluded that there is presence of danger in the scenario and hence the function for notification is invoked by the main program. As the considered security topic is sensitive, the framework defined also needs human intervention for exact decisions. But human resource is well utilized by having less interposing work. Rather is required to handle only when the system sends the notification.

Similarly all the frames in the captured video are analysed to detect suspicious activity and evaluated for the secureness of the real time scenario.

3.3 Workflow

1. Capturing input video and slicing into frames by various Matlab inbuilt functions such as NumberOfFrames, resd(VideoObj, frameno), imwrite() etc. Identifying reference background frame.
2. Traversing each frame of the video and analyzing with respect to the considered reference frame. If there is danger identified, then it is notified and paused for the response.
3. Adaptive background subtraction and noise removal by Image filtering methods available in Matlab and other predefined filters.
4. Blob detection and counting by Connected Area Component labeling technique. Break and start processing next frame if less than two people or blobs. Continue if more than 2 blobs detected.
5. Edge detection by means of Sobel edge detection technique for enhanced user interaction and human tracking.
6. Notifying through the relevant means if all of the conditions are satisfied: hand mould is matched, triangle like sharp object present in the frame and link existing between hand and object.

4. EXPERIMENTAL RESULT AND ANALYSIS

The proposed suspicious activity detection system is evaluated on real time recorded data set for validation of framework's procedural tasks. The video database includes 10 hours of video covering a wide variety of content. The format of reference video clips is 1280*1024 pixels and 30 frames/ sec. In our experiments, video clips that are taken as input are selected from reference dataset that are captured in 5-6 distinct locations.

The experiment conducted in three different environment as shown in Fig. 2 is taken for analysis. The first scenario is a well-lit classroom with benches, fans, curtains etc. as shown in fig. 2(a). The second scenario as shown in Fig 2(b) witnesses the experiment conducted in a university computer laboratory. The third case is a moderately-lit shuttered indoor space as shown in fig. 2 (c). Video clips shot in all the three locations is 30 to 45 seconds long per instance.



Figure 2. Experimental Input locations

The possible threat situation for the above three cases is shown in Fig. 3 First image depicts a scenario in which a person is trying to attack another person reading a book in a class, with a knife behind her back Fig 3(a). In the second image Fig 3(b) a person is pointing a knife towards the other which accounts for a threat situation. Fig 3(c) shows a person holding a knife in a shuttered indoor space, very close to another individual which might be a suspicious situation.

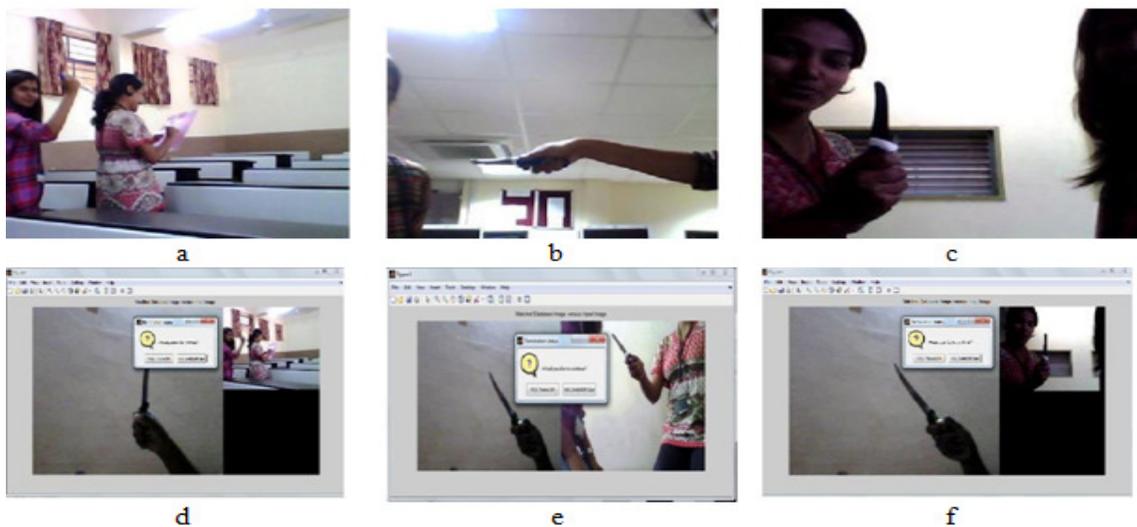


Figure 3. Threat Scenarios and their respective notifications

The possible threat in all the above three cases are notified by a pop up window shown in fig 3(d, e, f).

The entire process is done in sequence of steps. For the input image Fig 4(a) the different stages of the experiment is traced by the subsequent images in Fig 4. Fig 4 (b) shows the blob detection phase of the experiment where the two individuals are identified as two separate blobs. 4(b) is obtained by subtracting the Fig 4(a) with the reference background image with appropriate filters. The edge detection stage is shown by Fig 4(c) which is done using Sobel edge detection method with a defined threshold. The fourth image Fig 4(d) constantly tracks the foreground image in the video sequence. The tracking also helps in efficient implementation of user interface. For the danger notification which is the final step of the experiment, a pop-up window is displayed as shown in fig 4(e).

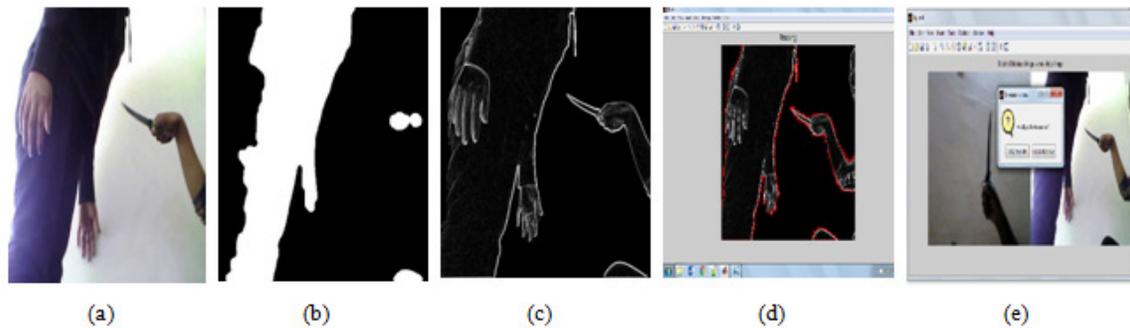


Figure 4. Different Stages of the experiment

Apart from all the true positives, the system also shows two false positives and three false negatives. For the shown fig 5(a) where the person is pointing a finger in the shape of a knife to a person in the vicinity gives a danger notification irrespective of the absence of the knife. The Hand shape is mistakenly seen as a knife by the system. In the second image shown in Fig 5(b) where the person's fist is falsely matched with the database image giving a false positive notification.



Figure 5. False Positives

The false negative cases is shown in Fig 6. In the first image Fig 6(a), no message is notified despite the presence of knife in one's hand. This is because the knife is being held in a different fashion for which system is not trained. Fig 6(b) also fails to notify any threat as the light intensity has caused shining glare on the knife which makes it undetected. In the third image Fig 2.5(c), the two blobs are overlapped along with the blob for the knife which makes it difficult for the system to detect.

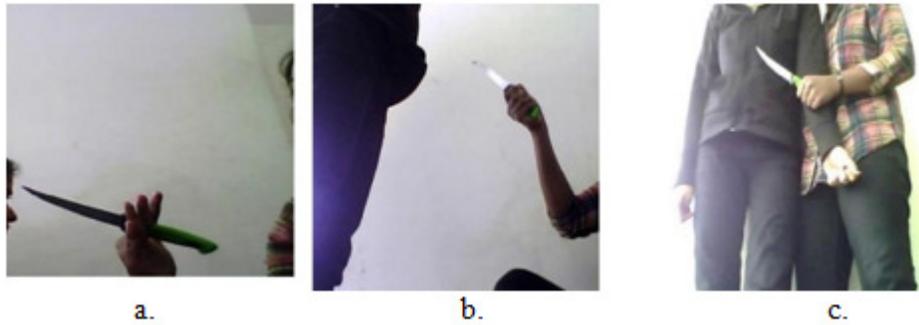


Figure 6. False Negatives

The system is evaluated by a precision and recall metric for the collected dataset from 3 different locations as mentioned in Fig 2.1. The particular rows represent the cases considered in Fig 2.2. It includes True positives, True negatives, false positives, false negatives count for the input video with 100 frames. Also Precision and Recall values are calculated. It is observed that the precision values decreases and recall increases with the subsequent test cases.

Table 1. PR Table of analysis result

Query Video	True Positives	True Negatives	False Negatives	True Negative	Precision (tp/(tp+fp))	Recall tp/(tp+fn)
1	61	0	30	9	1	0.67032967
2	41	10	12	37	0.803921569	0.773584906
3	65	25	7	3	0.722222222	0.902777778

5. CONCLUSION AND FUTURE WORK

This paper proposes a possible technique for Suspicious Activity Detection in indoor places to enhance the security system by improving the functionality of Surveillance camera to an intelligent device that is capable of detecting and notifying danger. An observer or the guard is relieved from the burden of continuous monitoring, may be physically or virtually watching enormous amount of video sequences captured by multiple Web cameras. Instead, intervention is serviced when the notification is sent. Due to its cost effectiveness, simple installation, scalability to different video resolutions, and once in a lifetime initialization, this is the feasible and practical solution to deploy in real scenarios.

The algorithm works efficiently in bright areas with 73 % accuracy, whereas functions moderately in less intensified areas with 67 % accuracy when experimented against the real time videos captured from distinct places.

Our goal is to rigorously continue to improve our detection and notifying system by adding various other features, such as implementation to detect other harmful activities like striking down, hitting, snatching the belongings and other physical abuse activities by supporting additional hand and object poses, expanding the dataset, experimenting in distinct possible locations, introducing machine learning techniques by using priori incidents knowledge with the focused concern of maintaining societal security. In the future, we expect promising results on

location specific distributional implementation with flattering performance. There is certainly room to improve the accuracy of the system to reduce the false positives and true negatives and to investigate on the category of activity and level of impact.

We believe, that the most rewarding impact is foreseen in security security however, the accuracy of the system needs to be enhanced for increased trustworthiness.

ACKNOWLEDGEMENT

We would like to thank Dr. Roopalakshmi, RVCE, and Bangalore for valuable discussions and many students of the University for creating vast test videos.

In addition, we would like to thank the anonymous reviewers for their valuable comments and suggestions that have helped to significantly improve the quality of this manuscript.

REFERENCE

- [1] Hidetomo Sakaino, (2013) "Video Based Tracking, Learning and Recognition Method for Multiple Moving Objects", IEEE transactions on circuits and systems for video technology, Vol. 23, NO. 10, pp 1661-1674.
- [2] Miwa Takai, (2010) "Detection of Suspicious Activity and Estimate of Risk from Human Behavior shot by Surveillance Camera", Second World Congress on Nature and Biologically Inspired Computing, pp 298-304, IEEE publishers.
- [3] Tao Luo, Ronald H. Y. Chung & K. P. Chow, (2014) "A Novel Object Segmentation Method for Silhouette Tracker in Video Surveillance Application", IEEE International Conference on Computational Science and Computational Intelligence, Vol. 1, pp 103-107.
- [4] Huang Li, Yihao zhang, Ming Yang, Yangyang Men & Hongyang Chao, (2014) "A rapid abnormal event detection method for surveillance video based on a novel feature in compressed domain of HEVC", IEEE International Conference on Multimedia and Expo(ICME),pp 1-6.
- [5] Lulu Chen , Hong Wei & James Ferryman, (2013) "A survey of human motion analysis using depth imagery", Pattern Recognition Letters, Vol 34,Issue 15, pp 1995-2006, Elsevier publishers.
- [6] Chengzhang Qu, Yuewei Lin, Dengyi Zhang & Song Wang, (2013) "Distant Human Interaction Detection from Kinect Videos", IEEE International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC), pp 1372 – 1375.
- [7] Manoranjan Paul, Shah M E Haque & Subrata Chakraborty, (2013) "Human detection in surveillance videos and its applications - a review", EURASIP Journal on Advances in Signal Processing, Springer Publication.
- [8] Garg R.,Aulakh I.K & Kumari N. , (2014), "A mathematical model to detect hand object from the scene", IEEE International Conference on Advance Computing (IACC), pp 1133-1136.
- [9] P Raghu Veera Chowdary, M Nagendra Babu, Thadigotla Venkata Subbareddy, Bommepalli Madhava Reddy & V Elamaran, (2014), "Image Processing Algorithms for Gesture Recognition using MATLAB" , IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), pp 1511-1514.

- [10] Mykyta Kovalenko, Svetlana Antoshchuk & Juergen Sieck , (2014), “Real-Time Hand Tracking and Gesture Recognition Using Semantic-Probabilistic Network”, 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, pp 269-274, IEEE Conf. Publications.
- [11] Pedro Cisneros & Paul Rodríguez, (2014) , “Practical Hand Tracking Solution by Alternating the Use of a priori Information”, 2014 IEEE 5th Latin American Symposium on Circuits and Systems (LASCAS), pp 1-4, IEEE Conf. Publications.
- [12] Javeria Farooq & Muhaddisa Barat Ali, (2014) “Real Time Hand Gesture Recognition for Computer Interaction”, 2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE), pp 73-77 ,IEEE Conf. Publications.
- [13] David Minnen & Zahoor Zafrulla , (2011), “Towards Robust Cross-User Hand Tracking and Shape Recognition”, 2011 IEEE International Conference on Computer Vision Workshops, pp 1235-1241.
- [14] Van-Toi Nguyen , Thi-Lan Le, Thanh-Hai Tran, R´emy Mullet & Vincent Courboulay , (2014) , “Hand posture recognition using Kernel Descriptor”, 6th International conference on Intelligent Human Computer Interaction, IHCI 2014,Procedia Computer Science, Vol. 39,pp 154-157 ,Elsevier Publications.
- [15] Radu-Daniel Vatavu, Ionut & Alexandru Zait, (2013) , “Automatic recognition of object size and shape via user-dependent measurements of the grasping hand”, International Journal of Human-Computer Studies ,Vol. 71,Issue 5,pp 590-607, Elsevier Publications.

AUTHORS

Nithya Shree R.

Students of R.V College of Engineering, Bangalore, India. I am interested in introducing technological enhanced solution for real time problems.



Rajeshwari Sah

Student of R.V College of Engineering, Bangalore, India. Most of my work are in the field of computer vision, image processing and machine learning domain



Shreyank N Gowda

Student of R.V College of Engineering, Bangalore, India. I am interested in the field of image processing, virtualization and cloud computing domains.

