VOL. 12, NO. 21, NOVEMBER 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

ENHANCEMENT OF SECURITY RELATED TO ATM INSTALLATIONS TO DETECT MISBEHAVIOR ACTIVITY OF UNKNOWN PERSON USING VIDEO ANALYTICS

K. Kajendran and A. Pravin
Department of Computer Science and Engineering, Sathyabama University, Chennai, Tamil Nadu, India
E-Mail: kajendran@yahoo.com

ABSTRACT

Almost everyone in the world use Automated Teller Machine (ATM) to either withdraw or deposit money from the bank account. However, most ATM Machines are prone to security risk, security breach, even though every ATM is equipped with CCTV installation. Some of security risk involves at ATM centre includes damaging ATM machine and decamping with cash, damaging CCTV camera to erase the identity of who involved in the crime, installing skimmer machine to capture debit card details and PIN number to create fake debit card, attacking ATM users and looting their valuable money withdrawn from ATM machine, murdering security guard for the purpose of looting ATM, withdrawing money by using fake ATM card. ATM crime is on the raise in the recent years due to its remote location. To enhance ATM security and to protect ATM centre form untoward incidents, new form of security framework needs to be identified. This paper presents a solution to identify unusual activities at ATM premises. For efficient detection of unusual activity in the ATM premises, this paper proposes a new mechanism of different window size to capture action rich frame that helps us to identify unusual activity and alert anticrime cell to avert ATM crime. This paper also proposes idea of recognizing unusual sounds such as breaking ATM machine with rod, screaming sound of customer, bursting sound of pistol within ATM centre which generates high decibels than decibels generated by normal human conversation and alert bank personal about ongoing crime that helps to catch culprit red handed. This mechanism allows the system to send notification message to bank authorities about ongoing troublesome activities, even CCTV camera damaged by criminal.

Keywords: ATM security, motion video detection, sound detection.

1. INTRODUCTION

An ATM is a part of our day to day activities. People using ATM keep on increasing. People are using ATM to withdraw their money without taking any help from bank employees. In India, in recent years, inauguration of new ATM by the bank in the remote location is happen on multiple folds to meet the customer demand. On the other side, assault on ATM users, murdering of ATM security guard, damaging ATM machine, decamping with looted money at the remote ATM is also on the raise. To safe guard ATM installation from looters, most of the ATMs in the remote location equipped with CCTV camera. Figure-1 shows a guard, who was sleeping inside an ATM, was beaten to death by looters in Jhunjhunu district, Rajasthan, India. In Figure-2 shows CCTV video footage, in which an attacker enter ATM booth and breaks ATM machine in Madurai, Tamil Nadu.

Figure-1. ATM guard beaten by looter.

Figure-2. Culprit attempt to damage ATM machine.

In Figure-3, a burglar attempt to loot cash from ATM by damaging ATM machine with hammer, at Mallapuram, Kerala, India. In Figure-4, CCTV footage shows man brutally attacking women with knife, inside an ATM in Bangalore, Karnataka, India, which left her right side paralyzed, skull fractured and most of her nose chopped off. The drawback of this CCTV installation is that a person has to monitor CCTV footage 24 hour per day and 7 days a week. This is tedious task. This mechanism has its own practical difficulties, because it is not possible for person to monitor the CCTV footage from all ATM center on the whole day the day, person might unconsciously sleep while monitoring CCTV footage. Damaging of CCTV camera by wrongdoer is also on the raise. CCTV footage only helps the anticrime cell and police to break through the crime incident and capture the

culprit earlier by analyzing video footage. Mostly it does not help to capture the offender on the spot, red handed. As of now, there is no mechanism to alert the concerned authorities, when untoward incident happen at ATM centre. We should need some mechanism to alert the person in the anticrime cell, in the nearby police station, or bank authority before culprit escape from the ATM premise. The second drawback is that if the CCTV camera was damaged by culprit before the crime was carried out then it is so difficult for investigator to breakthrough and identify the person who done the crime

Figure-3. Culprit attempt to damage ATM machine with hammer to loot money.

Figure-4. Man brutally attacking women with knife, inside an ATM.

This paper proposes an efficient solution to solve the above said problems. Our proposed system effectively detect unusual behaviors such as person with mask in a face, knife or pistol in a hand, or threaten somebody with gun, knife with in the ATM Center then send notification message to nearby police station or bank. This helps the police to catch culprit red handed within the ATM or near to the scene of crime. Usually no activities within the ATM generate high decibel sound. When unusual activities such damaging ATM machine and CCTV camera with hammer, crowbar, threatening a person with gun, screaming voice raised by the victim for help all will generate high decibel sound. This system also observes and measure the sound generated within ATM centre in decibel then the measured sound is above the normal level, this system sends SMS to nearby concerned authorities. This helps to notify police even CCTV camera was damaged by culprit

2. LITERATURE SURVEY

Here, we explore ATM security related research work carried out in the earlier period that helped us to propose this paper. Different approaches proposed by various researchers are analyzed to obtain best solution to our proposed system. Motion History Image (MHI) for identifying individual action was proposed by Atiqur Ahad *et al* [1]. Analysis of human movement by usage of temporal template was presented by Davis *et al* [2], Bobick *et al* [3]. Laptev *et al* [4, 5] suggested local spatial temporal feature with bag of word approach for achieving significant accuracy in detecting person action. Wang *et al* [6, 7] suggested that using of dense trajectories and dense sampling of point will get better the precision of detecting individual action. Dalal *et al* [8] suggested that combining histogram of oriented gradient (HOG) and Histogram of Optical Flow (HOF) helps to achieve better result. 3D gradient based descriptor suggested by Scovanner *et al* [10] through which orientated gradient can be measured using temporal dimension. Everts et al [11] and Souz *et al* [12] suggested that the inclusion of color helps better recognition of person action. Wang *et al* [7, 13] proposed that integration of HOB/HOF with dense trajectory descriptor will produce very good result compared with result of dense trajectory alone. Jain *et al* [14] suggested that incorporating differential motion sealer quantifier, divergence, curl, sheer feature to calculate descriptor achieved better accuracy. Kuehne *et al* [15] suggested histogram of gradient in which histogram is calculated by dividing frame into multiple blocks and producing normalized represent of image.

For machine to learn itself to detect human's unusual behavior, there are lot of classifier available. Compare with other classifiers support vector machine (SVM) produced better result to make machine to learn itself. [16] Nievas et al proved that uses of support vector machine in video surveillance achieve greater adaptive machine learning. SVM also support analysis of multiple classes. He suggests random forest for classification of video. Tuytelaars *et al* [9] suggested dense interest point for pattern recognition. Most of current approaches are not good enough to predict accurately on large and complex data set such as HMDB-52.These approaches has lot of performance issue for real time implementation. Requirements on the raise to build strong security framework for environment like ATM and similar premises.

3. PROPOSED SYSTEM

This proposed system consists of these modules.

a) Face mask detection and Denial of service
b) Detection of abnormal activity in ATM Center
c) Detection of unusual sound raised inside ATM premises

3.1 Face mask detection and Denial of service

In this module, the person attempting to withdraw money is examined to find whether he is wearing mask or helmet, before completing his ATM transaction. If so, service is denied for that person and the account from which he attempt to withdraw is blocked for next 24 hours. Notification SMS message saying that "somebody attempting to access your account and your account is block for 24 hours" is sent the mobile number of actual account holder.

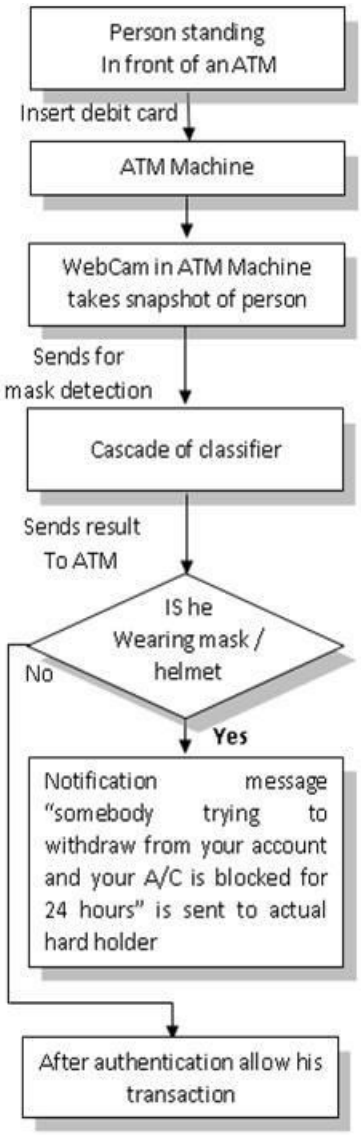

Figure-5. Processing of mask detection.

The detection of face without mask and with mask is done using cascade of classifier. The process of cascade of classifier is described next. For face detection cascade of classifier proposed by Paul Viola [18] is used, which offers one of the best face detection methods, compared with other techniques. Image repository that contains lot of images which includes both face without mask and face with mask which is used to train cascade of classifier. For feature extraction, this algorithm uses best threshold value for detection of both faces without mask and with mask. Features are selected based on minimum error rate for better classification of face from non-face images [20]. In this algorithm, the classifier is constructed with less number of significant features for quick classification. Initially each image given same weight and the frame is subdivided into two or more sub frames. After each successful classification, this algorithm increases the weight of misclassified image. This process is repeatedly done best accuracy or error rate is achieved. For efficient execution of algorithm, after every successful classification, frame with non-face region is eliminated. When applying classifiers in sequence, the rejected sub frames by the earlier classifier will not be processed by next classifiers. This system uses image representation referred as integral image introduced by Papageorgiou *et al*., is used, which facilitate rapid detection of face with mask [17].

3.2 Detection of abnormal activity in ATM Center

In this module, first generation of training dataset is done from collection of video. In this process the training video divided into frames. Every N number of successive frames is used to generate one concentrated frame called Action Rich Frame. For the calculation of magnitude of intensity of individual pixel in gray level with varying widow size, the root of sum of square is used. Window size N indicates number of frames that is used to generate feature rich frame, where N = 3, 5, 8, etc. Once matrix is generated from training video dataset, HOG is applied on action rich frame to obtain useful information from each new action rich frame. Once all videos are processed from training video repository, the training dataset was ready. Next, this module generates testing dataset from live video using similar procedure used in generation of training dataset. Once action rich frame was generated that feed into random forest classifier along with training data set, which detect whether action rich frame contain unusual human behavior

Figure-6. Processing steps involved in the proposed algorithm.

A step involved in the proposed algorithm is described next. For sequence of frames in video, the pixel intensity in every location of different frame will keep vary over time, when some activities going inside ATM. Calculation of magnitude of pixel intensity at specific location from sequence of frame is very useful information to identify different classes of activity. Action Rich Frame is usually generated from sequence of N frames. To calculate pixel intensity of each location of Action Rich Frame equation (A) is used

$$ARF(x, y) = \sqrt{(F_1^2 + F_2^2 + \ldots + F_N^2)} \quad (A)$$

Where
ARF - Action Rich Frame
F - Frames in the window set
N - Number of frame

Either three or five or eight consecutive frames (N=3, 5, 8) are used for generation new Action Rich Image or magnitude image. Table-1 shows an algorithm that describe step involved generating magnitude image using RSS.

Table-1. Proposed algorithm.

Algorithm for generation of Action Rich Frame
1) Take sequence of N frames from video feed
2) For each row *r* in all N frames For each column *c* in all N frames Compute ARF(r, c) = $\sqrt{F_1(r, c)^2 + F_2(r, c)^2 + \ldots + F_N(r, c)^2}$
3) Compute HOG
4) Predict unusual activity in Action Rich Frame computed in previous step
5) repeat Step 1 to Step 4 for next N frames

The Figure-6 shows processing step of algorithm in detail. Sequence of five frames picked from CCTV footage. Then RSS is applied on those five frames to generate Action Rich Frame. Then, to extract feature descriptor from action rich frame, HOG is applied on Action Rich Frame. The extracted feature descriptor is then feed into random forest classifier to predict unusual activity in the ATM premises. In HOG's feature based method, only region with motion within the frame is selected, and the extraction of motion information involves selection of interest point or area. So it produces better accuracy. But in ATM the area of CCTV coverage is very small, so that the entire frame would be classified without selection area of motion information. So we use HOG's direct method to extract motion information from entire image in the ATM premises

3.3 Detection of unusual sound within ATM

Sound raised in the ATM Centre was received through microphone, sampled at 44100Hz, stored as 16 bits quantized amplitude value and all possible sound spectrum components are recorded. Sounds such as gunshot, glass breaks generates highest frequency that shows a high energy content. Sound signal may contain characteristics such as temporal or spectral aspect, amplitude level (signal-to-noise ratio), duration, and time location as shown in Figure-7, which describes an amplitude level of long single human scream. Most of the impulsive sound shows mean duration of one second, but for scream, explosion, etc the mean time would reach 5 seconds. For example amplitude level of irregular human scream, burst, single gunshot are shown in Figures 8, 9, and 10.

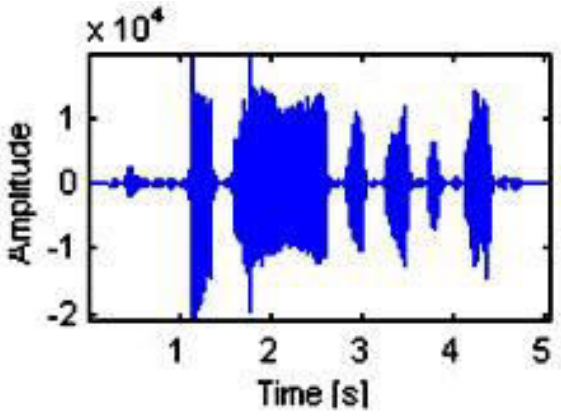

Figure-7. Amplitude level of long single human scream.

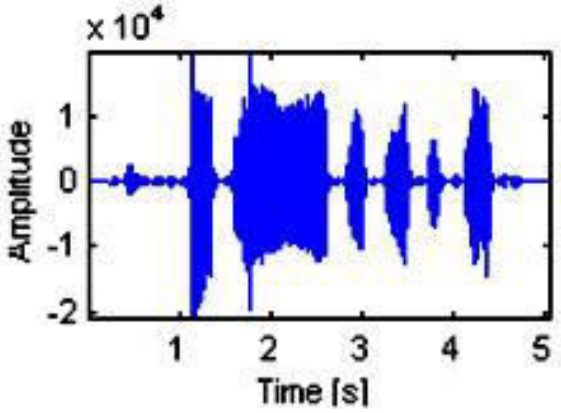

Figure-8. Amplitude level of long irregular human scream.

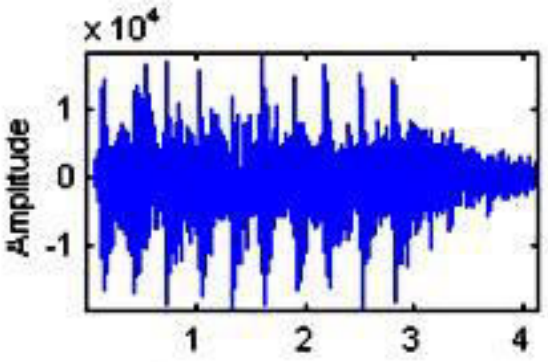

Figure-9. Amplitude level of Burst.

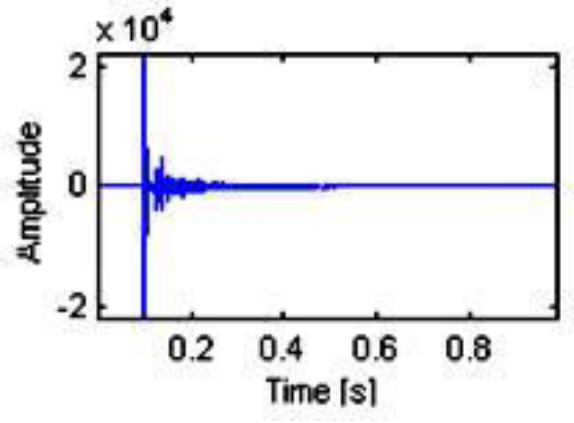

Figure-10. Amplitude level of Single gunshot.

Signal-to-Noise Ratio (SNR) is one of the best techniques for measurement of sound. The SNR can be calculated using the following formula (B):

$$\text{SNR} = 10 \log_{10} \frac{signal\ power\ Px}{noise\ poser\ Px}$$

$$= 10 \log_{10} \frac{\frac{1}{L} \sum_{I=1}^{L} Xi^2}{\frac{1}{N} \sum_{I=1}^{N} ni^2} \quad \text{(B)}$$

A unit of sound is measured in decibel (dB), which is used to calculate intensity of the sound. A normal conversation between humans can generate around 60 dB, whereas a gunshot or bursting of firecracker measured around 140 dB. Screaming and shouting of person measured around 90 dB.

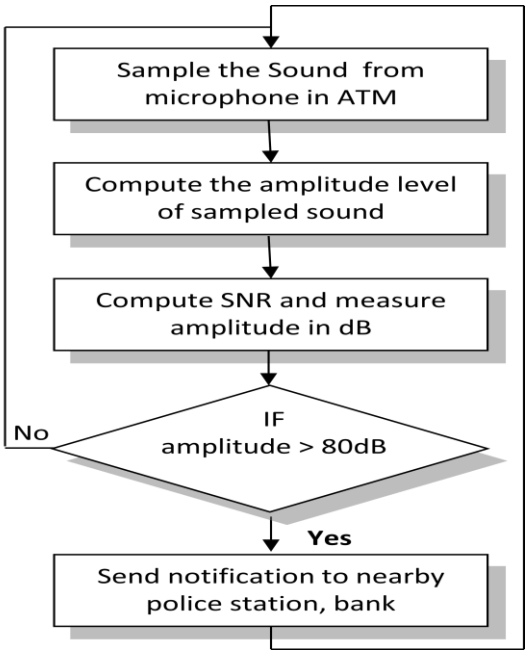

Figure-11. Process of unusual noise detection at ATM Centre.

In our proposed system, we observe that within ATM centre usually there was no sound except during

counting and dispatch of currency to customer by ATM Machine or some normal conversation between users. These actions usually generate sound that measured around 60dB to 65dB.

But unusual activities such as assault on security by offender, screaming voice raised by customer for help, breaking ATM machine with hammer, gunshot by culprit all will generate high decibel measured more than 80dB. Our proposed system observes and measure the sound generated within ATM centers in decibel and check whether the measured sound is above normal level. If so our proposed system sends notification to nearby bank and police station about unusual activities. This helps police to catch culprit red handed

To support sending SMS and recognize sound in ATM center from our application the following hardware were used: Arduino Uno microcontroller, GSM Modem, and Sound sensor. The Arduino Uno board (Spark fun DEV-09950) is a open source electronic platform that consists of 14 digital input and output pin and 6 analog input pins. To control an Arduino Uno board, program can developed on host computer to which board is connected through USB cable, by using simplified version of C/C++. An Open Arduino IDE is a open source development environment that can he used to write the program and dump into the Arduino board. Sketch for the Arduino Uno board can be created and uploaded to the board using an Open Arduino IDE. You can also map communication port to an Arduino board using an Open Arduino IDE, so that our program running in system can exchange message with an Arduino board through assigned communication port to complete some task such as sending SMS. An Arduino can take input from various sources like switches, sound sensors, variety of light, and other sources, GSM modem is used send SMS message through our program. Both GSM modem and sound sensor board is integrated with Arduino board, which is programmed with Arduino Software (IDE) that can take input from various sources like switches, sound sensors, variety of light, and other sources. Andunio configured in such a way that our program can take input and send output through communication port

Figure-12. Arduino Uno board.

Figure-13. Sound sensor.

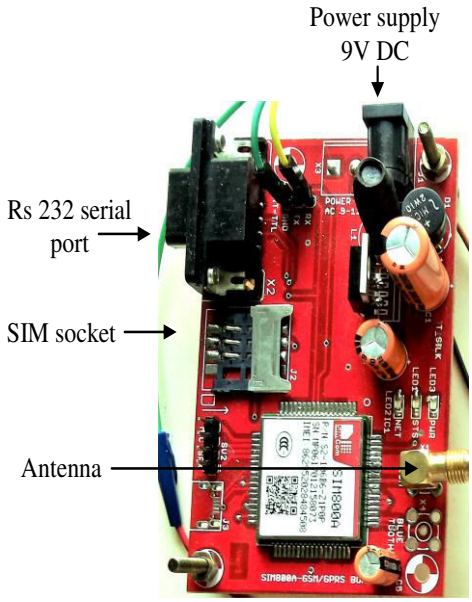

Figure-14. GSM modem.

5. CONCLUSIONS

Our proposed security framework for ATM premises offers three level of protection. The first level of protection is implemented through denial of service such as denying withdrawal of money for person wearing mask or helmet. In the second level of protection, if person armed with knife, pistol, firearm found within ATM premises then this system will alert the nearby police station with notification message describing about unusual activity of person. When troublesome incident happen within ATM center, this system will alert nearby police station, bank, which help the culprit caught red handed. Third, this system also employs sound recognition framework that capture the sound in the environment of ATM center and measure in decibel. If measured decibel is above the normal level, this system alert nearby police station. This helps to capture the culprit, even he dismantle

CCTV installed in ATM center. This system is tested with images from CAVIAR dataset, which produced best result. In future, recognition of unusual human activities in ATM can be detected with best level of accuracy by incorporating techniques like object detection. Our sound recognition system does not recognize voice, so in future voice recognition can also be included to enhance ATM security.

REFERENCES

[1] Atiqur Ahad *et al*. 2012. Motion history image and its variants and applications. Mach. Vision Appl. 23(2): 255-281.

[2] Davis *et al.* 1997. The representation and recognition of human movement using temporal templates, pp. 928-934. In: Computer Vision and Pattern Recognition, IEEE Proceedings.

[3] Bobick A. F., Davis J. W. 2001. The recognition of human movement using temporal templates: IEEE Transactions Pattern Analysis and Machine Intelligence. 23(3): 257-267.

[4] Laptev I. 2005. Space-time interest points. Int. J. Comput. Vision. 64(2-3): 107-123.

[5] Laptev I., Marszałek M., Schmid C., Rozenfeld B. 2008. Learning realistic human actions from movies. Proc. IEEE Conference Computer Vision and Pattern Recognition. pp. 1-8.

[6] Wang H., Klaser A., Schmid C., Liu C. L. 2011. Action recognition by dense trajectories. Proc. IEEE Conference Computer Vision and Pattern Recognition. pp. 3169-3176.

[7] Wang H., Klaser A., Schmid C., Liu C.L. 2013. Dense trajectories and motion boundary descriptors for action recognition. Int.J. Comput. Vision. 103(1): 60-79.

[8] Dalal N., Triggs B. 2005. Histograms of oriented gradients for human detection. Comput. Vision Pattern Recog. CVPR 1: 886-893.

[9] Tuytelaars T. 2010. Dense interest points: Proc. IEEE Conference Computer Vision and Pattern Recognition. pp. 2281-2288.

[10] Scovanner P., Saad A., Mubarak S. 2007. A 3-dimensional sifts descriptor and its application to action recognition. Proc. International conference on Multimedia. pp. 357-360.

[11] Everts I., Gemert J. C. van, Gevers T. 2013. Evaluation of color stips for human action recognition. Proc. IEEE Conference Computer Vision and Pattern Recognition. 2850-2857.

[12] Souza F., Valle E., Ca´mara-Cha´vez G., de Arau´ jo A. 2012. An evaluation on color invariant based local spatiotemporal features for action recognition. In IEEE SIBGRAPI.

[13] Wang H., Schmid C. 2013. Action Recognition with Improved Trajectories. Proc. IEEE International Conference in Computer Vision (ICCV). 3551-3558.

[14] Jain M., Je´gou H., Bouthemy P. 2013. Better exploiting motion for better action recognition. Proc. IEEE Conference Computer Vision and Pattern Recognition. 2555-2562.

[15] Kuehne H., Jhuang H., Stiefelhagen R., Serre T. 2011. HMDB: a large video database for human motion recognition. Proc. IEEE International Conference in computer vision (ICCV). 2556-2563.

[16] Nievas E.B., Suarez O.D., Garcı´a G.B., Sukthankar R. 2011. Violence detection in video using computer vision techniques, pp. 332–339. Computer Analysis of Images and Patterns. Springer, Berlin Heidelberg.

[17] C. Papageorgiou, M. Oren and T. Poggio. 998. A general framework for object detection, In: International Conference on Computer Vision.

[18] Paul Viola, Michael Jones: 2001. Rapid Object Detection using a Boosted Cascade of Simple Features, In Computer Vision and Pattern Recognition, IEEE Proceedings.

[19] Vikas Tripathi, Ankush Mittal: 2016. Real time security framework for detecting abnormal events at ATM installations, pp. 1-11 Journal of Real-Time Image Processing, Springer.

[20] O. Hori *et al*. 2006. Component-based robust face detection using AdaBoost and decision tree, 7th International Conference on Automatic Face and Gesture Recognition (FGR06).