Face Detection and Extraction from Low Resolution Surveillance Video Using Motion Segmentation

Vikram Mutneja[1]
Ph.D. Research Scholar,
I.K. Gujral Punjab Technical University, Kapurthala, Punjab (India)
vikram.mutneja@gmail.com

Dr. Satvir Singh[2],
Associate Professor,
I.K. Gujral Punjab Technical University Main Campus, Kapurthala, Punjab (India)
drsatvir.in@gmail.com

***Abstract*—Face detection is a prominent research domain in the field of digital image processing particularly in the field of video surveillance systems. Today is the world of video technology starting from low resolution videos to the high definition videos. The videos obtained from surveillance systems are often of low resolution due to the reasons such as distance between the camera and place of footage, environment factors, wide coverage area, installation problems, out of focus, bandwidth issue, hardware constraints, storage space limitations etc. because of which the frames need to be compressed or converted to lower resolution before storage. In this paper, we have worked on motion segmentation based face detection from low resolution surveillance videos. The motion segmentation is used to extract the region of interest from the current frame. Thereafter only the pixels obtained after the motion segmentation are subjected to the face detection process. The haar features based face detection has been used in this work, employing the image scaling to facilitate multi-scale face detection. Considerable search space reduction and efficiency boost has been achieved by proposed motion segmentation technique.**

***Keywords*-**Face Detection, Low Resolution Surveillance Videos, Motion Segmentation, Haar Features, AdaBoost

I. INTRODUCTION

Video surveillance systems also known as CCTV (Closed circuit television) are the systems which use video camera for the purpose of surveillance. These systems keep on recording the video footage of the scene continuously 24 hours a day and keep storage of video as per the storage capacity of the system e.g. of last one week or one month. Video surveillance is in use since quite a long time for the monitoring and security purposes in various public and private places such as railway stations, offices, banks, roads, showrooms and shopping malls. The method of monitoring had been manual most of the times which usually is prone to human errors on account of factors such as fatigue, lack of human attention. The advent of technology and computational power has given way to transformation of traditional manual surveillance systems into intelligent video surveillance systems, which not only record the data, but also do the intelligent i.e. automatic monitoring of the video. Intelligent surveillance video systems aim at two main functions. Firstly detection of the objects of interest e.g. people, vehicles. Secondly tracking, activity analysis and recognition of the objects for the events detection and recognition e.g. detection and tracking of human face and head for the analysis of human attention, detection of a "person smoking" activity which involves detection of human, hands, cigarette and smoking activity.

There is great thrust in research in the field of automatic surveillance video systems. The proposed work is concerning detection and extraction of human faces from the low resolution surveillance video sequence. In the surveillance video systems where task is to detect, track and recognize people as well as analyze people activities, detection and extraction of human faces is of paramount importance. It is very important to attach the identity to persons being detected and tracked in the video. From the fact that human faces are used as biometric entity, human faces are generally used to attach identity to a detected human in the surveillance video. Detecting the human faces in surveillance videos is a challenging task on account of various factors such as illumination, low resolution of surveillance cameras, human pose, face pose, facial gestures, head pose, face occlusions such as goggles, scarf, face hair and various face and head accessories. The facial images in surveillance videos are of very low resolution, therefore it is challenging task to detect the faces from low resolution surveillance videos.

The training and detection framework used in this work has been derived from our previous work related to modified haar features and AdaBoost based face detection system Mutneja and Singh (2017). This paper is structured as follows: Section 2 provides Literature Survey, Section 3 gives details of proposed technique for the face detection algorithm, Section 4 explains experimental setup, Section 5 discusses experimental results, and finally Section 6 discusses conclusions and future Scope.

II. LITERATURE SURVEY

Sarkar et al. (2012) worked on multiple face detection and tracking from low resolution video sequence, they used skin color information for face region estimation, eyes and mouth region localization in the detected skin region to confirm it as face region. Chen et al. (2007) used video object and skin color segmentation for face localization and neural networks for face quality analysis. Kasturi et al. (2009) proposed very robust framework for performance evaluation of face detection and tracking in surveillance videos. Zhu and Ramanan (2012) presented a unified model for face detection, pose estimation and landmarks localization in real world cluttered images. Wang (2014) proposed a complete algorithmic description, a learning code and a learned face detector that can be applied to any color image. Since the Viola-Jones algorithm typically gives multiple detections, a post-processing step is also proposed to reduce detection redundancy using a robustness argument.

Zakaria and Suandi (2011) used combination of neural network and adaboost, Huang et al. (2011) used combination of Genetic algorithm and neural network, Martinez-Gonzalez and Ayala-Ramirez (2011) used neural networks for real time face detection, Jaisakthi and Aravindan (2011), used data and sensor fusion technique using SVM, Guan et al. (2012) proposed face localization using fuzzy classifier, haar features and YcbCr color features, Pan et al. (2013) used combination of haar like, local binary patterns and speeded up robust features is conjunction with SVM and PSO for multi-view face detection, Hiremath et al. (2012) implemented fuzzy geometric face model for searching face region using prominent face features such as eyes and mouth to detect faces. Seyedarabi et al. (2009) used skin color and face edge information to develop a fuzzy rule based classifier to extract head candidate from image using YcbCr colour space model. Ming Ouhyoung et al. (2012)used real time depth sensors for nose detection for human face localization to overcome face occulsions. Kuo et al. (2010) used fuzzy c-means for color recognition of objects in surveillance videos.

During the past two decades of research in the field of face detection, Viola and Jones (2004) did seminal work in face detection, key contributions of their work were: new image representation called integral image to facilitate faster calculation of features, adaboost learning algorithm for classifiers, cascade classifiers for faster computation. As per surveyed by Belaroussi and Milgram (2012), growing research field is concentrating in developing appearance based models for multi-view and rotation invariant face detections. In case of color image sequence, using skin color results in faster face localization and poses estimation.

Alionte and Lazar (2015) proposes practical implementation of a face detector based on Viola-Jones algorithm using Matlab cascade object detector is presented. Employing the system type object vision.CascadeObjectDetector, eight face detectors were developed using the trainCascadeObjectDetector function and tuning the number of cascade layer and the False Alarm Rate. For different tuning parameters, the performances of the face detectors were analyzed.

III. PROPOSED ALGORITHM

The face detection process in proposed work is composed of multiple steps. Firstly the difference of current video frame is found with respect to the previous frame to find the motion segmented pixels. Further only the motion segmented pixels are subjected to the generation of sub-images, to be parsed for the face classification. The size of the sub-images is same as the detector window. The detector has been trained using the example images of size 18×18. The configuration of the trained detector being used is as in table 1.

TABLE 1: DETECTOR CONFIGURATION

Training Images Size	18×18
Haar Features Pool Size	32384
Number of Weak Classifiers:	1147
Number of Stages in Cascade	19
Configuration	[2, 2, 3, 5, 5, 10, 20, 30, 40, 50, 60, 80, 90, 100, 110, 120, 130, 140, 150]

The multi-scale face detection has been facilitated by applying image scaling. The minimum and maximum values of scaling factors are calculated based upon the size of input image, minimum size of face to be detected and maximum image size which can be handled by system. Algorithm 1 shows the working of the face detection process and Algorithm 2 shows the function to process the sub-images generated from the motion segmented pixels called from the previous algorithm.

Algorithm 1 Face Detection Process

1: Load the First Frame of the Video ($pFrame$), Size of the Video Frame, $NR \times NC$
2: Compute the
   $LowestScalingFactor(LR)$,
   $UpperScalingfactor(UR)$,
   $N = NumberofScalingFactors$
3: Load the Haar Classifiers of the Detection Cascade Kernel
4: Initialize and load the function (ProcessMotion) for performing the detection on motion segmented pixels
5: Loop1: Load the Current Frame of Video ($cFrame$)
6: Extract the motion segmented pixels as vectors MX and MY, which are x,y co-ordinates respectively of motion segmented region
7: Initialize the Array to Hold the Locations of Detected Faces
8: Loop2: For multi-scale face detection, i=1:N
9: Scale Image and Motion Pixel Vectors XM, YM: $img1 = Img * i * LR, XMi = XM * i * LR, YMi = YM * i * LR$
10: Determine $xMax$, $yMax$ Maximum pixel locations for sub-image generation based upon size of scaled image, $xMin = yMin = 2$.
11: Eliminate the the values from XMi and YMi which are out of range $xMintoxMax$ and $yMintoyMax$
12: Initialize the array to hold the sub-images to be generated.
13: Loop3: To scan for pixel locations from XMi and YMi for sub-images generation with shift interval $SL(i)$
14: **if** X, Y lies in the previous detected face area **then**
15: Skip Rest and Go to Next Motion Pixel Location (Loop3)
16: Create the sub-image, compute its integral image and add it to array of sub-images (Integral Versions)
17: Iterate Loop2 for Next Motion Pixel Location
18: Call the function (Pass the argument: Generated array of sub-images (Integral Versions) to detect faces from them)
19: Update Array of Detected Faces for Scaling Factor=$i * LR$, Iterate Loop2 for Next Scaling Factor
20: All Faces Detected in Current Frame,
21: Set $pFrame = cFrame$, i.e. Previous frame=Current frame and Iterate Loop1 for Next Frame

IV. EXPERIMENTAL SETUP

The machine running windows 8.1 (64 Bit) on Intel core i3 1.9 GHz has been used to test the proposed method. The work has been done in MATLAB version 8.2.0.701 (R2013b).

V. RESULTS AND DISCUSSION

The proposed algorithm has been tested on the test videos from the low resolution surveillance systems from the dataset INRIA (2004) with specifications: Frame Width: 384 pixels, Frame Height: 288 pixels, Frame Rate: 25 Frames per Second, Data Rate: 1184 kbps, Total Bitrate: 1184 kbps. Figure 1 and Figure 2 shows the result of motion segmentation on few of the frames of test videos from dataset INRIA.

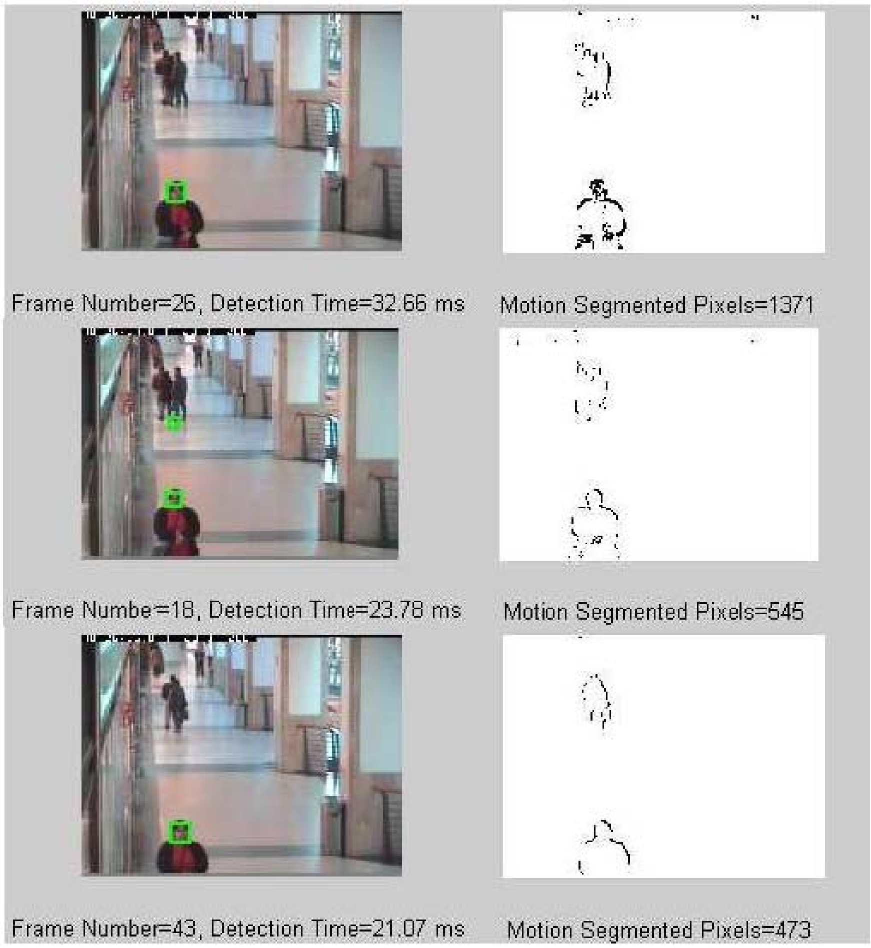

Figure 1: Result Motion Segmentation on a Test Video (OneLeaveShop1cor.mpg) From Dataset INRIA

TABLE 2: PROCESSING TIME AND NUMBER OF SEGMENTED PIXELS OF STARTING 20 FRAMES (TEST VIDEO 1)

Frame	Detection Time (ms)	Segmented Pixels
1	0.00	0
2	85.24	315
3	22.65	252
4	20.68	368
5	22.59	529
6	22.27	632
7	27.79	736
8	23.64	663
9	21.74	546
10	18.47	531
11	19.50	485
12	21.16	568
13	21.76	577
14	23.20	677
15	0.00	0
16	33.50	1387
17	24.81	725
18	26.51	893
19	24.30	669
20	22.53	552

Table 2 shows the time cost and number of pixels segmented by proposed motion segmentation method on a test video1 (“OneLeaveShopReenter2cor.mpg” from INRIA). We have been able to achieve the speed of processing of the order of 37.86 fps, with the average detection rate of 98.86% at the false acceptance rate 4.44 %.

Figure 2: Detection Results on Few Test Video Frames

Table 3 shows the time cost and number of pixels segmented on a test video 2(“OneShopOneWait1cor.mpg” from INRIA) by proposed motion segmentation method. We have been able to achieve the speed of processing on the test video 2 of the order of 54.57 fps, with the average detection rate of 99.52% at the false acceptance rate 4.21 %. From the results, it is inferred that we have been able to achieve the promising results in terms of processing speed as well as detection accuracy on the test videos. Further by comparison of results of first and second video, it is inferred that better time efficiency has been achieved in the second video because of less number of motion segmented pixels to be processed.

Algorithm 2 Function to Process Sub-Images for Face Detection Process (ProcessMotion)

Load Variables/Function Arguments $DetectorWindowSize(DWS)$, $NumberofWeakClassifiers(NCF)$, $NumberofStagesofCascadeKernel(numLevels)$, $ArrayofFeatures(F)$, $Thresholds(TH)$ and $NumberofSub-Images(numSubImages)$
Loop1: for Iterating the Sub-images, For sIdx=1:$numSubImages$
Compute Absolute Base Address for pixel Operations: $BaseAddress = sIdx * DWS * DWS$
Loop2: for Feature Number $(fNum)$, for fNum=1:NCF
Initialize array to hold binary results of decision stumps, $face\_or\_noface(NCF)$
Fetching the co-ordinates of sub-rectangles from feature array and compute Haar Feature Value
Compare it for Lower and Upper Bounds of Thresholds
Update $face\_or\_noface$ Array
Iterate Loop2 for evaluating all the Features
Loop3: For i=1:$numLevels$; Operation of Detection Cascade
Threshold value $V(i)$ computed of Stage number i
if $if(V(i) > TH(i))$ **then**
Go to Loop3, to Iterate to Next Stage of Detection Cascade
else
$status(sIdx) = 0$, Detected as Non_Face **return**
if $i == numLevels$ **then** $status(sIdx) = 1$ (Detected as Face) **return**
Iterate Loop1 for Next Sub-image

TABLE 3: PROCESSING TIME AND NUMBER OF SEGMENTED PIXELS OF STARTING 20 FRAMES (TEST VIDEO 2)

Frame	Detection Time (ms)	Segmented Pixels
1	0.00	0
2	81.51	25
3	18.37	52
4	14.04	51
5	13.64	21
6	13.82	40
7	13.92	19
8	14.25	41
9	14.33	33
10	14.97	19
11	14.46	31
12	14.58	17
13	15.35	28
14	13.60	24
15	14.67	10
16	13.36	27
17	13.56	15
18	15.19	53
19	14.58	28
20	14.38	8

VI. CONCLUSION AND FUTURE SCOPE

In the proposed system, we have developed the algorithm for the detection and extraction of faces from low resolution surveillance videos using motion segmentation based technique. The inter-frame difference is performed and the pixels having the difference above the set threshold level are marked as motion segmented pixels, which are subjected to generation of sub-images for the detection of faces. The detector of small size i.e. 18×18 has been used so as to target the low resolution faces. The testing has been performed on the test videos from the dataset INRIA. From the results achieved, we contend that proposed algorithm is very effective in detection of faces from low resolution surveillance videos.

We intend to do further improvements in the proposed algorithm by incorporating the handling of faces with severe head orientations and occlusions. We further want to integrate the proposed algorithm with the complete video surveillance based face biometric system and acceleration with the help of GPU computing.

ACKNOWLEDGMENT

The proposed work has been carried under the research related to first author's Ph.D. in field of facial image processing from low resolution surveillance videos, registered as part time research scholar under I.K. Gujral Punjab Technical University, Kapurthala, Punjab (India).

REFERENCES

[1] Elena Alionte and Corneliu Lazar. A practical implementation of face detection by using matlab cascade object detector. In System Theory, Control and Computing (ICSTCC), 2015 19th International Conference on, pages 785–790. IEEE, 2015.

[2] Rachid Belaroussi and Maurice Milgram. A comparative study on face detection and tracking algorithms. Expert systems with Applications, 39 (8): 7158–7164, 2012.

[3] Tse-Wei Chen, Shou-Chieh Hsu, and Shao-Yi Chien. Automatic feature-based face scoring in surveillance systems. In Multimedia, 2007. ISM 2007. Ninth IEEE International Symposium on, pages 139–146. IEEE, 2007.

[4] Chen-Ning Guan, Chia-Feng Juang, and Guo-Cyuan Chen. Face localization using fuzzy classifier with wavelet-localized focus color features and shape features. Digital Signal Processing, 22 (6): 961–970, 2012.

[5] PS Hiremath, Manjunath Hiremath, and R Mahesh. Face detection and tracking in video sequence using fuzzy geometric face model and motion estimation. International Journal of Computer Applications, 58 (15), 2012.

[6] Chen-rong Huang, Jia-li Tang, and Yi-jun Liu. A new face detection method with ga-bp neural network. In Wireless Communications, Networking and Mobile Computing (WiCOM), 2011 7th International Conference on, pages 1–4. IEEE, 2011.

[7] EC Funded CAVIAR project/IST 2001 37540 INRIA. 2004. URL http://homepages.inf.ed.ac.uk/rbf/CAVIAR/.

[8] SM Jaisakthi and Chandrabose Aravindan. Face detection using data and sensor fusion techniques. In Soft Computing and Pattern Recognition (SoCPaR), 2011 International Conference of, pages 274–279. IEEE, 2011.

[9] Rangachar Kasturi, Dmitry Goldgof, Padmanabhan Soundararajan, Vasant Manohar, John Garofolo, Rachel Bowers, Matthew Boonstra, Valentina Korzhova, and Jing Zhang. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol. IEEE Transactions on Pattern Analysis and Machine Intelligence, 31 (2): 319–336, 2009.

[10] Jong Yih Kuo, Tai Yu Lai, Fu-Chu Huang, and Kevin Liu. The color recognition of objects of survey and implementation on real-time video surveillance. In Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on, pages 3741–3748. IEEE, 2010.

[11] Angel Noe Martinez-Gonzalez and Victor Ayala-Ramirez. Real time face detection using neural networks. In Artificial Intelligence (MICAI), 2011 10th Mexican International Conference on, pages 144–149. IEEE, 2011.

[12] Ming Ouhyoung, Hong-Shiang Lin, Yi-Ting Wu, Yi-Shan Cheng, and Dominik Seifert. Unconventional approaches for facial animation and tracking. In SIGGRAPH Asia 2012 Technical Briefs, page 24. ACM, 2012.

[13] Hong Pan, Yaping Zhu, and Liangzheng Xia. Efficient and accurate face detection using heterogeneous feature descriptors and feature selection. Computer Vision and Image Understanding, 117 (1): 12–28, 2013.

[14] Rajib Sarkar, Sambit Bakshi, and Pankaj K Sa. A real-time model for multiple human face tracking from low-resolution surveillance videos. Procedia Technology, 6: 1004–1010, 2012.

[15] Hadi Seyedarabi, Saeed Mahdizadeh Bakhshmand, and Sohrab Khanmohammadi. Multi-pose head tracking using colour and edge features fuzzy aggregation for driver assistant system. In Signal and Image Processing Applications (ICSIPA), 2009 IEEE International Conference on, pages 385–390. IEEE, 2009.

[16] Paul Viola and Michael J Jones. Robust real-time face detection. International journal of computer vision, 57 (2): 137–154, 2004.

[17] Yi-Qing Wang. An analysis of the viola-jones face detection algorithm. Image Processing On Line, 4: 128–148, 2014.

[18] Zulhadi Zakaria and Shahrel A Suandi. Face detection using combination of neural network and adaboost. In TENCON 2011-2011 IEEE Region 10 Conference, pages 335–338. IEEE, 2011.

[19] Xiangxin Zhu and Deva Ramanan. Face detection, pose estimation, and landmark localization in the wild. In Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, pages 2879–2886. IEEE, 2012.

[20] Vikram Mutneja and Satvir Singh. Modified viola–jones algorithm with gpu accelerated training and parallelized skin color filtering-based face detection. Journal of Real-Time Image Processing, pages 1–21, 2017.

AUTHORS PROFILE

Mr. Vikram Mutneja is Ph.D. research scholar in I.K. Gujral Punjab Technical University, Kapurthala (Punajb, India) in the discipline Electronics Engineering. He received his Bachelor's degree (B.Tech.) from Guru Nanak Dev University, Amritsar, Punjab (India) with specialization in Electronics & Communication Engineering in year 1998, Master's degree (M.Tech.) from Giani Zail Singh College of Engineering & Technology, Bathinda (Punjab) (India) under Punjab Technical University (PTU, Jalandhar) with first division in Electronics & Communication Engineering in year 2008. He is a Life Member of ISTE. He has industrial experience of around 5 years (1998-2003) in which he worked in diverse fields such as Hardware and Networking Technical Support, Software Development for Web and Embedded Systems. During his teaching experience of around 13 years (2003-2016) he worked mainly in the areas of embedded systems, VLSI and Digital Signal Processing. He is working presently as Assistant Professor in department of Electronics & Communication Engineering in Shaheed Bhagat Singh State Technical Campus, Ferozepur, Punjab (India). His current field of research includes areas of image processing for detection and extraction of facial features from low resolution surveillance videos, and acceleration using GPU computing. (Vikram.mutneja@gmail.com)

Dr. Satvir Singh is Associate Professor in I.K. Gujral Punjab Technical University (Main Campus), Kapurthala (Punajb, India). He received his Bachelor's degree (B.Tech.) from Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, Punjab (India) with specialization in Electronics & Communication Engineering in year 1998, Master's degree (M.E.) from Delhi Technological University (Formerly, Delhi College of Engineering), Delhi (India) with distinction in Electronics & Communication Engineering in year 2000 and Doctoral degree (Ph.D.) from Maharshi Dayanand University, Rohtak, Haryana (India) in year 2011. He is a Life Member of ISTE and Corporate Member of IETE. During his 15 years of teaching experience he served as Assistant Professor and Head, Department of Electronics & Communication Engineering at BRCM College of Engineering & Technology, Bahal, (Bhiwani) Haryana, India and as Professor & Head, Department of Electronics & Communication Engineering at Shaheed Bhagat Singh State Technical Campus (Formerly, SBS College of Engineering & Technology), Ferozepur Punjab, India. His fields of special interest include Evolutionary Algorithms, High Performance Computing, Type-1 & Type-2 Fuzzy Logic Systems, Wireless Sensor Networks and Artificial Neural Networks for solving engineering problems. (Drsatvir.in@gmail.com)