A Survey on Face Detection Techniques

Vikram Mutneja[1] and Satvir Singh[2]
[1,2]*Department of Electronics & Communication Engineering,*
Shaheed Bhagat Singh State Technical Campus, Ferozepur
E-mail: [1]Vikram.mutneja@gmail.com, [2]DrSatvir.in@gmail.com

***Abstract*—A lot of work has been done till date in the area of face detection. Work on detection of faces has taken leap in various directions since its inception early in the area of image processing and computer vision. The earlier work intended to be focused on detecting faces from still images. With the advent of videos, the focus shifted to detecting and extracting faces from videos. In recent time, with the advent in the surveillance video systems i.e. smart CCTVs, a lot of work has been done in extracting faces from surveillance videos. The main challenges in this area are accuracy and speed. Because of low resolution and small size of faces from surveillance video systems, the accuracy is greatly affected. Further, also the processing of video requires working on high fps (frames per second) videos which affects the speed of operation. So in this paper, we have surveyed the recent work done in the area of detecting and extracting the faces from still images as well as videos, particularly surveillance videos.**

Keywords: Face Detection, Surveillance Video, 3-D Features, CENTRIST

I. INTRODUCTION

Face detection is not a straightforward problem because of factors such as pose variation, occlusions, image orientation, illumination conditions and facial expressions. Face detection techniques can be classified mainly into four categories: Knowledge based methods, Feature-Invariant approaches, Appearance based methods and Template matching methods [1]. Knowledge based methods use the face knowledge to encode rules such as a face often appears in an image with two eyes that are symmetric to each other, a nose and a mouth. Challenge is that it is difficult to translate human knowledge into well defined rules. Feature-invariant approaches extract features such as edges, facial features such as eyes, nose, mouth, and hair line and a statistical model is built to describe their relationships. Challenges in this area are such as lighting, shadows etc. Appearance based methods extract features based upon appearance such as Eigen faces-PCA, Neural Networks, SVM, Ada Boost. In template matching based methods predefined templates have to be stored and correlation values with the standard patterns are computed e.g.: for the face contour, eyes, nose and mouth independently. Limitations so far is that it cannot effectively deal with variation in scale, pose and shape. Challenges are how to represent the template, how to model deformations, efficient matching algorithms etc..

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, pose variation, facial expressions, face occlusions. The facial images detected from surveillance videos are of very low resolution and not suitable to be applied to face recognition system. Therefore there is also the need to estimate and enhance the quality of detected facial images to bridge the gap between face detection and face recognition systems.

To target the challenges in the area of detecting and extracting facial images from videos, many approaches have been suggested such as incorporating the motion and skin color cues [2] [3], mainly to reduce the search area. Video based techniques work on motion estimation, video object segmentation, background subtraction, skin color detection, object tracking based methods. A recent trend in face detection is to combine multiple information sources such as color, motion, contour etc. More the number of information channels, more will be the accuracy of the system but at the cost of increased detection time.

Inter frame difference technique is most simple and efficient, given (Eq. 1).

$$\Delta(n)_{(x,y)} = I(n)_{(x,y)} - I(n-1)_{(x,y)} \qquad (1)$$

where n represents time and (x, y) the pixel location. The current frame difference $\Delta(n)_{(x,\,y)}$ is compared with a threshold value, in case it is more than threshold value, the corresponding pixel location is set to 1 otherwise 0. In case of color image, the calculation has to be performed in each color space separately and then aggregated to create final binary image showing the moving parts. Challenges in this area are such as noise removal, threshold estimation etc.

Another category of techniques in motion estimation are optical flow based methods. The biggest advantage of optical flow based techniques is that they can be used even if camera is moving or the background is changing fast. In case of cluttered environment, where background is complex and changing fast, background subtraction technique cannot be used. However optical flow technique has shortcomings as it is more complex, time consuming and poor anti-noise performance.

Skin color cue has also been used in color images to reduce the search area. Difference efficiency has been achieved with various color formats for skin color extraction. RGB format has been found to be best for human vision but not for skin color detection. YCbCr format has been found to give best performance for the skin color detection, as it concentrates the skin colored pixels in a small intensity range. Another challenge in this area is that different cameras produce different colors of same object in scene. So there is also the need of color correction techniques to color of images from different sources.

II. Literature Survey

A lot of work in recent years has been done to develop and use hybrid features in boosting based face detection algorithms. Jun, Bongjin *et al.,* [4] proposed two novel local transform features: Local Gradient Patterns (LGP), modified version of LBP (local Binary Patterns) and Binary Histogram of Oriented Gradients (BHOG), modified version of HOG (Histogram of Oriented Graphics), which were proved to be Faster in computation as compared to LBP and HOG. They proposed hybrid feature that combines various local transform features including LBP, LGP and BHOG by means of AdaBoost method for face and human detection, Hybridization results in improvement of detection performance on account of LBP's robustness to global illumination variations, LGP's robustness to local intensity changes and BHOG's to local pose changes. The proposed local transform features and its hybrid feature have been found to be effective for face and human detection in terms of performance and operating speed.

Some work in recent years has been done to use holistic representation of features for detection, enhancement and recognition of faces. The idea is that the main aim of extracting faces from surveillance videos is to identify the person by applying face recognition. So system efficiency is expected to improve if holistic features are used instead of using separate features for each task such as detection, enhancement and recognition of faces. Bharadwaj, S. *et al.,* [5] have Studied the possibility of using Holistic descriptors Gist and HOG to use in biometric quality assessment of facial images. The spatial properties are preserved in representation, called as Gist. The promising results were obtained in use of above features for quality assessment in face biometrics.

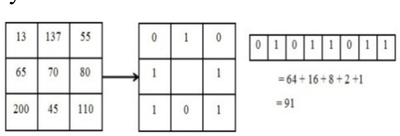

Fig. 1 Calculation of 2-D LBP Features

Another dimension in which the work of extracting faces from videos has been addressed is the use of volumetric features i.e., 3-D version of 2-D features such as HAAR, LBP (see Fig. 1), HOG features. The recent work by Martinez-Diaz, Yoanna, *et al.,*. [6] has been done in the use of spatio-temporal based features EVLBP (Extended Volumetric Local Binary Patterns) (see Fig. 2) for detecting the faces from videos. The motivation behind this work was enhancement of efficiency of the system by using 3-D features, which can encode N number of frames in its generation. The main challenge in this area is to select the number of frames, which have to be processed at a time to encode the features. More the value of N, more will be the speed of processing, but accuracy may reduce as it may encode the non facial information while feature encoding. The selection of value of N depends upon how fast the contents of scene in video are changing or how fast the object is moving. Use of EVLBP has been found to give better performance than spatial LBP.

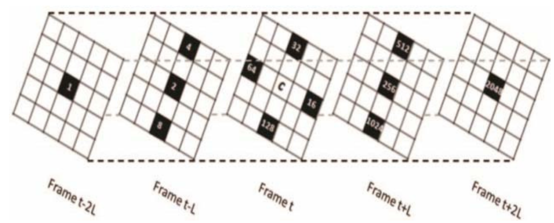

Fig. 2 3-D LBP Features (EVLBP)

Another dimension in which the work of face detection has leapt is the use of parallel and GPU computing [7] [8] [9] [10] [11] [12] to enhance the speed of operation at various stages. The GPU computing has been explored in accelerating the training of boosting based face classifiers. Oro, David, *et al.* [11] presented techniques to increase the performance of the cascade evaluation kernel, which is the most resource-intensive part of the face detection pipeline. Worked on handling problem of GPU underutilization, and achieved a 5X speedup in 1080p videos on average over the fastest known implementations, while slightly improving the accuracy. Also studied the parallelization of the cascade training process and its scalability under SMP platforms. The proposed parallelization strategy exploits both task and data-level parallelism and achieves a 3.5x speedup over single-threaded implementations.

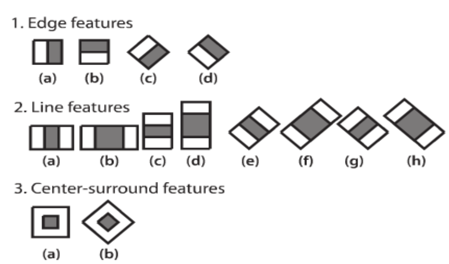

Fig. 3 HAAR Features (Basic & Extended)

The recent work by Jiamin wu *et al.* [13] have proposed new methodology, with the name C^4 for object detection, which has performed very well on face detection too. They have been able to achieve the 20 fps speed and state of the art detection efficiency. They have used and applied the conjecture that contours and signs of comparisons with the neighboring pixels are key information for object detection. The C^4 means, Contour, Cascade Classifier and CENTRIST visual descriptor [14]. Authors have proposed the future scope of work in accelerating the speed of operation of C^4.

They have used new visual descriptor CENTRIST [14] which has been found to be suitable match for contour based object detection. CENTRIST visual descriptor encodes the signs of neighboring comparisons. It has been derived from Census Transform (CT) which was originally designed to establish correspondence between neighboring patches [15]. Please see Fig. 4, it shows the calculation of CT value for the center pixel. The CT image C of an input image I is generated by replacing a pixel with its CT value. The CENTRIST descriptor is a histogram with 256 bins, which is a histogram of these CT values in an entire image or a rectangular region in an image.

$$\begin{array}{c|c|c} 32 & 64 & 96 \\ \hline 32 & \mathbf{64} & 96 \\ \hline 32 & 32 & 96 \end{array} \Rightarrow \begin{array}{ccc} 1 & 1 & 0 \\ 1 & & 0 \\ 1 & 1 & 0 \end{array} \Rightarrow (11010110)_2 \Rightarrow \mathrm{CT} = 214$$

Fig. 4 Census Transform

El-Sayed *et al.* [16] have used mean of medians of CbCr color correction approach to enhance the combined SMQT (Successive Mean Quantization transform) features and SNoW (Sparse Network of Winnow) classifiers (SFSC). The proposed method has been found to be more efficient and accurate as compared to original SFSC method.

Viola Jones *et al.,* [17] did seminal work in face detection, implemented state-of-the-art face detector. This algorithm is still continuing to be leader in modern face detection implementations such as in mobile devices, also inbuilt in OPENCV and MATLAB. Key contributions of their work were: Scale invariance, New image representation called integral image to facilitate faster calculation of Haar features, Uses Machine learning, Adaboost learning algorithm for combining the weak classifier to obtain strong classifiers, Cascade of haar classifiers (see Fig. 3) for faster computation. It was 15 times faster than previous work, can be generalized to detect any type of object. There were some limitations such as the training set of negative examples has to be small to make training feasible, the process of creating the detector cascade is based on trial and error process, training process is lengthy (may take few days, depends upon speed of machine), can handle up to limited rotation angle of faces (± 15 degrees in plane, ±45 degrees out of plane), fails in face occlusions and low brightness of faces. See Fig. 5, how Haar

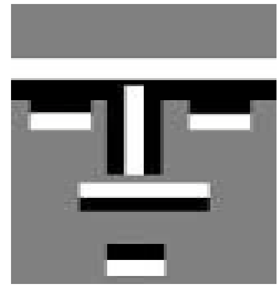

Fig. 4 Applying HAAR Features on Face

features are able to search for various face features such as eyes, nose, lips etc.

Nasrollahi, K. *et al.,* [18] did excellent work on generating good quality frontal face image from low resolution video sequence, used viola and jones [2] face detector, face quality is estimated based upon facial features: sharpness, brightness, resolution, head Pose. They used auto associative memories for the head pose estimation. Further generated high resolution frontal face image using reconstruction and learning based super resolution techniques in cascade.

Bagdanov, A. *et al.* [19] worked on multi-face detection, tracking, facial image quality analysis and face-log generation. They developed multipose face detector, based on Adaboost face detector, used lateral and frontal face detectors. System has been evaluated on 10 hours of realistic surveillance videos, with both quantitative and qualitative analysis. However it reported to have limitations such as the proposed face-logging system is appropriate for situations in which face size is bounded, illumination conditions are consistent with the images used to train the Adaboost detectors in their multipose face detectors.

Chen, Tse-Wei *et al.* [20] combined the spirit of image based face detection and essence of video object segmentation to filter out face candidates. Developed a face scoring technique, using eight scoring functions based on feature extraction technique, used a single layer neural network training system to obtain an optimal linear combination to select high quality faces. The face detector was based on skin color detection and video object segmentation. Scoring functions used eight functions: Skin color coverage, Luminance variation, Circularity measurement, Eye-pixel Histogram, Ratio, Angle, Symmetry and Hair. All eight functions were combined using fuzzy logic to calculate the final score.

Chang-yeon, Jo. *et al.* [21] worked on LBP, face images are divided into M small non-overlapping regions, LBP histogram are extracted from each sub-region, All such histograms are combined together into a single spatially enhanced feature histogram, Extracted feature describes local as well as global shape of the face images, a variant of AdaBoost, Gentle AdaBoost has been used to select the features and train the classifier. Cascade of classifiers is used for enhancing the performance. The developed algorithm has been found to computationally efficient and tested on Mobile platform.

III. Conclusion and Future Scope

There is no doubt that lot of research work has been done in the area of face detection but the goal is still far from achieved: To mimic the human vision of detecting and identifying the human faces. So to meet that goal, still a lot of work has to be done in this area. As per literature survey, following directions for future work in this area are being proposed:

1. The training of Haar features in seminar viola jones' face detector takes a long time, which may be couple of days if used serial processing. There is scope of work to apply the parallel computing to enhance the speed of features training. Till date not much work has addressed the performance
2. Comparisons of various software platforms such as MATLAB, use of GPU in C/C++ environment, use of GPU in MATLAB environment. So there is scope of using optimization work to address the issue of speed of training of features.
3. In the use of volumetric features, there is open research area in: a) Integrating the descriptor with the scanning strategy, b) Setting criteria for selecting the optimal number of frames to encode the descriptor, c) Investigating in using same feature space for face detection & recognition.
4. Use of holistic features for performing various tasks in the process of face extraction from video such as face detection, face quality estimation, face quality enhancement and face recognition instead of using separate feature for each task.
5. Using motion information in creating face-logs from the video.

References

[1] C. Zhang and Z. Zhang, “A survey of recent advances in face detection,” Tech. rep., Microsoft Research, Tech. Rep., 2010.

[2] H.M.M.R. Hiremath, P.S., “Face detection and tracking in video sequence using fuzzy geometric face model and motion estimation,” International Journal of Computer Applications, Vol. 15, No. 58, pp. 12–16, 2012.

[3] H. Seyedarabi, S.M. Bakhshmand, and S. Khanmohammadi, “Multipose 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. IEEE, 2009, pp. 385–390.

[4] B. Jun, I. Choi, and D. Kim, “Local transform features and hybridization for accurate face and human detection,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 35, No. 6, pp. 1423–1436, 2013.

[5] S. Bharadwaj, M. Vatsa, and R. Singh, “Can holistic representations be used for face biometric quality assessment?” in Image Processing (ICIP), 2013 20th IEEE International Conference on. IEEE, 2013, pp. 2792–2796.

[6] Y. Martinez-Diaz, H. Mendez-Vazquez, N. Hern´andez, and E. GarcIa-Reyes, “Improving faces/ non-faces discrimination in video sequences by using a local spatio-temporal representation,” in Biometrics (ICB), 2013 International Conference on. IEEE, 2013, pp. 1–5.

[7] B. Bilgic, B.K. Horn, and I. Masaki, “Efficient integral image computation on the gpu,” in Intelligent Vehicles Symposium (IV), 2010 IEEE. IEEE, 2010, pp. 528–533.

[8] H. Jia, Y. Zhang, W. Wang, and J. Xu, “Accelerating viola-jones facce detection algorithm on gpus,” in High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems (HPCC-ICESS), 2012 IEEE 14th International Conference on. IEEE, 2012, pp. 396–403.

[9] J. Kong and Y. Deng, “Gpu accelerated face detection,” in Intelligent Control and Information Processing (ICICIP), 2010 International Conference on. IEEE, 2010, pp. 584–588.

[10] E. Li, B. Wang, L. Yang, Y.T. Peng, Y. Du, Y. Zhang, and Y.J. Chiu, “Gpu and cpu cooperative accelaration for face detection on modern processors,” in Multimedia and Expo (ICME), 2012 IEEE International Conference on. IEEE, 2012, pp. 769–775.

[11] D. Oro, C. Fern’ndez, C. Segura, X. Martorell, and J. Hernando, “Accelerating boosting-based face detection on gpus,” in Parallel Processing (ICPP), 2012 41st International Conference on. IEEE, 2012, pp. 309–318.

[12] G. Wei and C. Ming, “The face detection system based on gpu+ cpu desktop cluster,” in Multimedia Technology (ICMT), 2011 International Conference on. IEEE, 2011, pp. 3735–3738.

[13] J. Wu, N. Liu, C. Geyer, and J.M. Rehg, “C¡ sup¿ 4¡/sup¿: A real-time object detection framework,” 2013.

[14] J. Wu and J.M. Rehg, “Centrist: A visual descriptor for scene categorization,” Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 33, No. 8, pp. 1489–1501, 2011.

[15] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in Computer Vision XECCV’ 94. Springer, 1994, pp. 151–158.

[16] M.A. El-Sayed and N.G. Ahmed, “Enhanced face detection technique based on color correction approach and smqt features,” Journal of Software Engineering and Applications, Vol. 6, pp. 519, 2013.

[17] P. Viola and M.J. Jones, “Robust real-time face detection,” International Journal of Computer Vision, Vol. 57, No. 2, pp. 137–154, 2004.

[18] K. Nasrollahi and T.B. Moeslund, “Extracting a good quality frontal face image from a low-resolution video sequence,” Circuits and Systems for Video Technology, IEEE Transactions on, Vol. 21, No. 10, pp. 1353–1362, 2011.

[19] A. Bagdanov, A. Del Bimbo, F. Dini, G. Lisanti, and I. Masi, “Compact and efficient posterity logging of face imagery for video surveillance,” IEEE Multimedia, Vol. 19, No. 4, pp. 48–59, 2012.

[20] T.W. Chen, S.C. Hsu, and S.Y. Chien, “Automatic feature-based face scoring in surveillance systems,” in Multimedia, 2007. ISM 2007. Ninth IEEE International Symposium on. IEEE, 2007, pp. 139–146.

[21] Chang-yeon, “Face detection using lbp features,” 2008.