A Survey On Face Detection and Recognition

Neha Rathore, Deepti Chaubey[*2], Nupur Rajput[#3]

***Abstract*— recently face recognition and face detection is attracting much attention in the society of network multimedia information access. Areas such as network security, content indexing and retrieval, video compression benefits from face recognition technology because "people" are the centre of attention in lot of video. We are making recce of different papers that were published recently and a comparison between them and by that making a survey.**

***Keywords*— Face Recognition; Biometric Identification; Eigen Faces; Network Security and Surveillance; still image; video sequence.**

I. Introduction

A facial recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame from a video source. One of the ways to do this is by comparing selected facial features from the image and a facial database. In today's networked world, the need to maintain the security of information or physical property is becoming both increasingly important and difficult. From time to time we hear about the crimes of credit card fraud, computer break-ins by hackers, a security breaches in company or government building. The face recognition technology is based in a field called 'biometrics. Biometric access control are automated methods of verifying or recognizing the identity of a living person on the basis of some physiological characteristics, such as figure prints or facial features, or some aspects of the person's behaviour, like his/her handwriting style. Since biometric systems identify a person by biological characteristics, they are difficult to forge [1].

With the rapid increase of computational powers and availability of modern sensing, analysis and rendering equipment and technologies, computers are becoming more and more intelligent. Many research projects and commercial products have demonstrated the capability for a computer to interact with human in a natural way by looking at people through cameras, listening to people through microphones, understanding these inputs, and reacting to people in a friendly manner.

One of the fundamental techniques that enable such natural human-computer interaction (HCI) is face detection.

Face detection is the step stone to all facial analysis algorithms, including face alignment, face modelling, face relighting, face recognition, face verification/authentication, head pose tracking, facial expression tracking/recognition, Gender/age recognition, and many more. Only when computers can understand face well will they begin to truly understand people's thoughts and intentions.

Face detection is anonymous by definition. The digital image is not matched against a database of known individuals. In contrast to face detection is the science of facial recognition. This software solution concentrates on making a positive identification of the individual against a database that archives personal information. Confidence factor is a key metric to avoid improper identification.

"Face Recognition" is a very active area in the Computer Vision and Biometrics fields, as it has been studied vigorously for 25 years and is finally producing applications in security, robotics, human-computer-interfaces, digital cameras, games and entertainment.

"Face Recognition" generally involves two stages:

A. *Face Detection*, where a photo is searched to find any face (shown here as a green rectangle), then image processing cleans up the facial image for easier recognition.
B. *Face Recognition*, where that detected and processed face is compared to a database of known faces, to decide who that person is (shown here as red text).

It is usually harder to detect a person's face when they are viewed from the side or at an angle, and sometimes this requires 3D Head Pose Estimation. It can also be very difficult to detect a person's face if the photo is not very bright, or if part of the face brighter than another or has shadows or is blurry or wearing glasses, etc.

However, Face Recognition is much less reliable than Face Detection, generally 30-70% accurate. Face Recognition has been a strong field of research since the 1990s, but is still far from reliable, and more techniques are being invented each year.

Eigenfaces (also called "Principal Component Analysis" or PCA), a simple and popular method of 2D Face Recognition from a photo, as opposed to other common methods such as Neural Networks or Fisher Faces. Eigen faces are a set of Eigen vectors used in computer vision problem of human face recognition [2].

Network security and Surveillance for face recognition is based on monitoring of the behaviours, activities, or other changing information, usually of people for the purpose of influencing and managing, directing, or protecting.

II. A LOCAL FEATURE BASED FRAMEWORK

In recent advances in a project being undertaken to trial and develop advanced surveillance systems for public safety is proposed that is, local facial feature based framework for both still image and video-based face recognition.

After the bombing attack in 2005, special attentions have been paid to the use of CCTV for surveillance to prevent such attacks in the future. Based on the number of CCTV cameras on Putney High Street, it is “guesstimated” [3] that there are around 500,000 CCTV cameras in the London area and 4,000,000 cameras in the UK. This implies that there is approximately one camera for every 14 people in the UK. Given the huge number of cameras, it is impossible to hire enough security guards to constantly monitor all camera feeds. Hence, generally the CCTV feeds are recorded without monitoring, and the videos are mainly used for a forensic or reactive response to crime and terrorism after it has happened. However, the immense cost of successful terrorist attacks in public spaces shows that forensic analysis of videos after the event is simply not an adequate response. In the case of suicide attacks, there is no possibility of prosecution after the event, so only recording surveillance video provides no terrorism deterrent. There is an emerging need to detect events and persons of interest from CCTV videos before any serious attack happens. This means that cameras must be monitored at all times.

However, two main constraints restrict human monitoring of the CCTV videos. One important issue is the limitation of the number of videos that a person can monitor simultaneously. For large amount of cameras, it requires a lot of people resulting in high ongoing costs. Another issue is that such a personnel intensive system may not be reliable due to the attention span of humans decreasing rapidly when performing such tedious tasks for long time.

One possible solution is advanced surveillance systems that employ computers to monitor all video feeds and deliver the alerts to human operators for response. Because of this, there has been an urgent need in both the industry and the research community to develop advanced surveillance 2 EURASIP Journal on Image and Video Processing systems, sometimes dubbed as Intelligent CCTV (ICCTV). In particular, developing total solutions for protecting critical infrastructure has been on the forefront of R&D activities in this field [4].

III. THE VIOLA-JONES FACE DETECTOR

If one were asked to name a single face detection algorithm that has the most impact in the 2000’s, it will most likely be the seminal work by Viola and Jones [5]. The Viola-Jones face detector contains three main ideas that make it possible to build a successful face detector that can run *in real time*: the integral image, classifier learning with Ada Boost, and the intentional cascade structure.

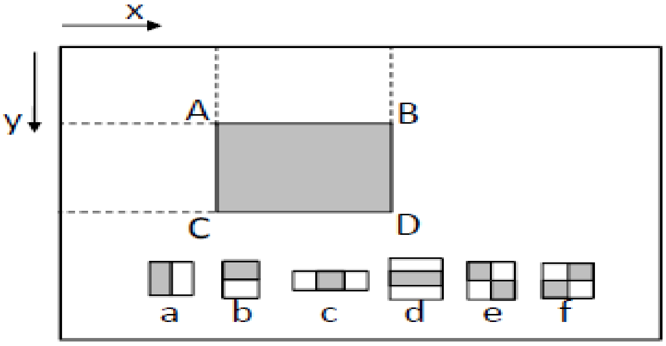


Fig. Illustration of integral image and Haar-like Rectangle features.

A. *The internal image*

Integral image, also known as a summed area table, is an algorithm for quickly and efficiently computing the sum of values in a rectangle subset of a grid. It was first introduced to the computer graphics field by Crow [6] for use in mipmaps. Viola and Jones applied the integral image for rapid computation of Haar-like features, as detailed below:

The integral image is constructed as followes:

$$ii(x, y) = \sum_{x'<x, y'<y} i(x', y')$$

Where *ii*(*x; y*) is the integral image at pixel location (*x; y*) and *i*(x_0; y_0) is the original image.

B. *Ada Boost learning*

Boosting is a method of finding a highly accurate hypothesis by combining many “weak” hypotheses, each with moderate accuracy. For an introduction on boosting, we refer the readers to [7] and [8].

The AdaBoost (Adaptive Boosting) algorithm is generally considered as the first step towards more practical boosting algorithms [9, 10]. In this section, following [11] and [12], we briefly present a generalized version of AdaBoost algorithm, usually referred as *Real Boost*. It has been advocated in various works that Real Boost yields better performance than the original AdaBoost algorithm.

IV. FEATURE EXTRACTION

As mentioned earlier, thanks to the rapid expansion in storage and computation resources, appearance based methods have dominated the recent advances in face detection.

The general practice is to collect a large set of face and non-face examples, and adopt certain machine learning algorithms to learn a face model to perform classification. There are two key issues in this process: what features to extract, and which learning algorithm to apply. In this section, we first review the recent advances in feature extraction. The Haar-like rectangular features are very efficient to compute due to the integral image technique, and provide good performance for building frontal face detectors. In a number of follow-up works, researchers extended the straightforward features with more variations in the ways rectangle features are combined.

For instance, as shown in Fig. 5, Lienhart and Maydt[13] generalized the feature set of [14] by introducing 45 degree rotated rectangular features (a-d), and center-surround features (e-f). In order to compute the 45 degree rotated rectangular features, a new rotated summed area table was introduced as:

$$rii(x,y) = \sum_{x' \le x, |y-y'| \le x-x'} i(x',y')$$

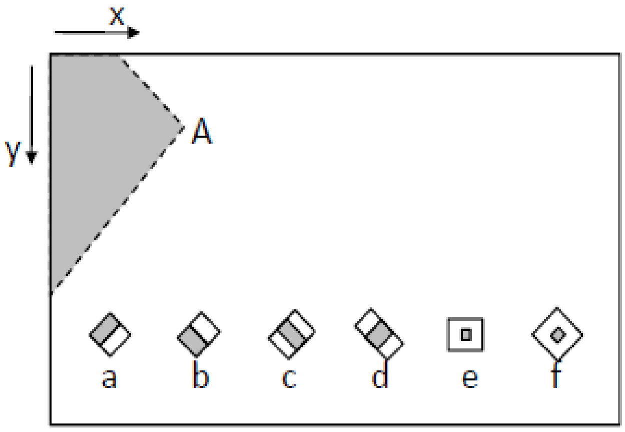


Fig. The rotated integral image/ summed area table.

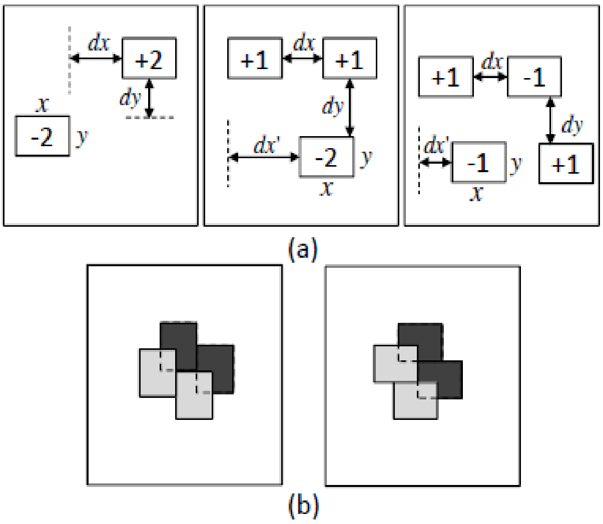


Fig. (a.) Regular features with flexible sizes and distances introduced in [15] and (b.) Diagonal filters in [16]

A number of researchers noted the limitation of the original Haar-like feature set in [17] for multi-view face detection, and proposed to extend the feature set by allowing more flexible combination of rectangular regions. For instance, in [18], three types of features were defined in the detection sub-window, the rectangles are of flexible sizes *x* £ *y* and they are at certain distances of (*dx; dy*) apart. The authors argued that these feature scan be non-symmetrical to cater to non-symmetrical characteristics of non-frontal faces. Jones and Viola [19] also proposed a similar feature called diagonal filters. These diagonal filters can be computed with 16 array references to the integral image.

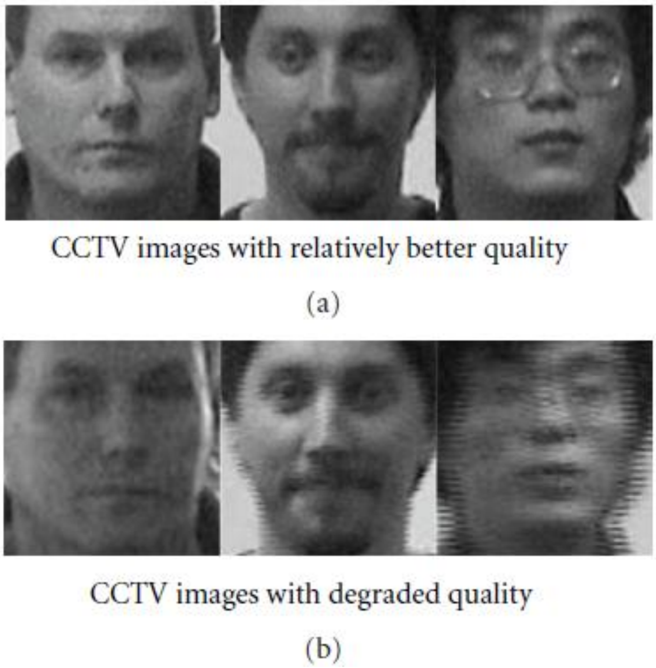
V. STILL IMAGE VERSUS VEDIO SEQUENCE

For face recognition in surveillance scenarios, identifying a person captured on image or video is one of the key tasks. This implies matching faces on both still images and video sequences. It can be further classified into three categories: still image to still image matching, video sequence to video sequence matching, and still image to video sequence matching.

Automatic face recognition for still images with high quality can achieve satisfactory performance, but for video-based face recognition it is hard to attain similar levels of performance. Compared to still images face recognition, there are several disadvantages of video sequences. First, images captured by CCTV cameras are generally of poor quality. The noise level is higher, and images may be blurred due to movement or the subject being out of focus. Second, image resolution is normally lower for video sequences. If the subject is very far from the camera, the actual face image resolution can be as low as 64 by 64 pixels. Last, face image variations, such as illumination, expression, pose, occlusion, and motion, are more serious in video sequences. These effects are illustrated in Figure 3. Images in the first row are CCTV images with relatively good quality. The second row shows degraded images, where the left-hand side picture shows the effect of out of focus, the middle picture displays the effect of interlacing due to object movement and the right-hand side one illustrates the combination of out of focus and interlacing. To comparison with the still image shown in Figure, it can be seen that the image quality of CCTV cameras (even high-end ones) is much worse than still images. In addition, the poor quality, low resolution, and large variation will result in uncertainty of the face detector, which is the first important step of any automatic face recognition system. Faces extracted from poor-quality video scan have higher false detection rate and larger alignment errors, which may have great influence on the performance [19].

However, there are some major advantages of video sequences. First, we can employ spatial and temporal information of faces in the video sequence to improve still recognition performance.

Second, psychophysical and neural studies have shown that dynamic information is very crucial in the human face recognition process. Third, with redundant information, we can reconstruct more complex representations of faces such as 3D face model or super-resolution images and apply them to improve recognition performance. Fourth, some online learning techniques can be applied for video-based face recognition to update the model over time. Since we need to do both still image and video-based face recognition under surveillance conditions; the above approaches are not suitable. Most still image face recognition techniques are not appropriate for surveillance images due to the following concurrent and uncontrolled factors. The pose, illumination, and expression variations are shown to have great impact on face recognition [20]. Image resolution change due to variable distances to cameras is another factor that influences the recognition performance. The face localization error induced by automatic face detector will definitely affect the recognition results as there are no guarantees that the localization is perfect.



CCTV images with relatively better quality

(a)

CCTV images with degraded quality

(b)

Fig. Normalized video of face image captured by CCTV cameras.

VI. Enhancement of MRH for Video-based Face Recognition

For intelligent surveillance systems, automatic face recognition should be performed for both still images and video sequences. Thus, normal video-based face recognition techniques are not suitable for this task since they are designed only for video-to-video matching. In an attempt to retain the ability for still image face recognition and to be capable for still-to-video and video-to-video matching, we propose the following approaches to enhance MRH for face recognition on videos. In this section, we explore four methods that operate on features to build up a more representative model for classification as well as four methods that operate on distance between vectors to improve the performance. By investigating these approaches, we attempt choose a best suitable method that takes advantage of multi frame information in a computationally inexpensive manner for image set and video-set matching. As part of the investigation into this problem, a subset of LFW database is used for image set matching, test and a large-scale audiovisual database called “Mobio Biometry” (MOBIO) [21] is used for video-set matching, respectively.

A. Operation on Feature.

In this approach, several methods are inspected, which utilize multiple feature vectors of the sample images in a set to build up a more representative model of faces. In other words, they attempt to extract more meaningful new features from the existing features. In the following sections, we will discuss them in more detail.

A.1. Feature Averaging: To extend still image face recognition for video sequences, a direct approach is applying still image recognition for each frame in the video set. But this approach is computationally expensive and does not fully utilize spatial and temporal information of the video. Given an example, to identify a face from a probe video with *f* frames in a video database with *V* video sequences, the thorough search needs to perform the still image matching by $V \times m \times v$ times, where v is the average frames per sequence. Generally, for only a 10-second video, it would contain about 300 frames (with a normal frame rate at 30 fps). This means that the calculation for video is about 90000 times of that for still image.

Inspired by a recent paper published in Science [22], we propose the following approach by averaging MRH facial features. Different from [23], where a simple image averaging is applied, we average the features due to the observation that image averaging is only helpful for holistic facial features and impairs the local facial features. Assume that MRH is applied on frame *k* of video *p* to extract the histogram *hpk* then the final description of the face in this video is by averaging as follows:

$$h_p = \frac{1}{v_p}\sum_{k=1}^{v_p} h_{p,k},$$

VII. CONCLUSION

In this paper, we surveyed some of the recent advances in face detection. face detection in completely unconstrained settings remains a very challenging task, particularly due to the significant pose and lighting variations. In our in-house tests, the state-of-theart face detectors can achieve about 50-70% detection rate, with about 0.5-3% of the detected faces being false positives. Consequently, we believe there are still a lot of works that can be done to further improve the performance.

The most straightforward future direction is to further improve the learning algorithm and features. The Haar features used in the work by Viola and Jones are very simple and effective for frontal face detection, but they are less ideal for faces at arbitrary poses. Complex features may increase the computational complexity, though they can be used in the form of a post-filter and still be efficient, which may significantly improve the detector's performance.

However, other learning algorithms such as SVM or convolution neural networks can often perform equally well, with built-in mechanisms for new feature generation. The modern face detectors are mostly appearance based methods, which means that they need training data to learn the classifiers. Unsupervised or semi-supervised learning schemes would be very ideal to reduce the amount of work needed for data collection. In environments which have low variations, adaptation could bring very significant improvements to face detection.

There are number of promising techniques to give better results and can be use for effective face recognition and detection in still and video images.

REFERENCES

[1] Facial Recognition Application.

[2] Face Recognition with Eigenface Servo magazine tutorial and sourcecode.

[3] L. M. Fuentes and S. A. Velastin, "From tracking to advanced surveillance," in Proceedings of the International Conference on Image Processing (ICIP '03), pp. 121–124, September 2003.

[4] F. Ziliani, S. Velastin, F. Porikli et al., "Performance evaluation of event detection solutions: the CREDS experience," in Proceedings of IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS '05), pp. 201–206, September 2005.

[5] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. of CVPR, 2001.

[6] F. Crow. Summed-area tables for texture mapping. In Proc. of SIGGRAPH, volume 18, pages 207–212, 1984.

[7] R. Meir and G. R¨atsch. An introduction to boosting and leveraging. S. Mendelson and A. J. Smola Ed., Advanced Lectures on Machine Learning, Springer-Verlag Berlin Heidelberg, pages 118–183, 2003.

[8] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Dept. of Statistics, Stanford University, 1998.

[9] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In European Conf. on Computational Learning Theory, 1994.

[10] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences, 55(1):119–139, 1997.

[11] R. E. Schapire and Y. Singer. Improved boosting algorithms using confidence-rated predictions. Machine Learning, 37:297–336, 1999

[12] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: a statistical view of boosting. Technical report, Dept. of Statistics, Stanford University, 1998

[13] R. Lienhart and J. Maydt. An extended set of Haar-like Features for rapid object detection. In Proc. of ICIP, 2002.

[14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. of CVPR, 2001.

[15] S. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum. Statistical learning of multi-view face detection. In Proc. Of ECCV, 2002.

[16] M. Jones and P. Viola. Fast multi-view face detection.Technical report, Mitsubishi Electric Research Laboratories, TR2003-96, 2003

[17] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Proc. of CVPR, 2001.

[18] S. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum. Statistical learning of multi-view face detection. In Proc. Of ECCV, 2002.

[19] V. Jain and E. Learned-Miller. FDDB: A benchmark for face detection in unconstrained settings. Technical report, University of Massachusetts, Amherst, 2010.

[20] D. Keren, M. Osadchy, and C. Gotsman. Antifaces: A novel fast method for image detection. IEEE Trans. on PAMI, 23(7):747–761, 2001.

[21] S.Marcel, C. McCool, P. M. Ahonen et al., "Mobile biometry mobio) face and speaker verification evaluation," in Proceedings of the 20th International Conference on Pattern Recognition, 2010.

[22] R. Jenkins and A. M. Burton, "100% Accuracy in automatic face recognition," Science, vol. 319, no. 5862, p. 435, 2008
.