Summary of Research on Face Detection and Tracking Technology

Xiaodan Lu[1, a]

[1.]Zhengzhou University of Industry Technology, Department of Information Engineering, Zhengzhou, China

[a.]Houyanyang521@sina.com

Keywords: Face detection; Face tracking; Application; Research status

Abstract. Face detection and tracking is a hot topic in the field of computer vision. This paper introduces the problem in detail. Firstly, the difficulty of face detection and face tracking is analyzed; Secondly, the wide application of face detection and tracking technology in video surveillance, human-computer interaction and information security is introduced; Finally, the research status of face detection is analyzed. On this basis, the research status of long-term and long-term human face tracking in any environment is analyzed step by step, and the TLD tracking framework is introduced in detail.

Introduction

People in the image tend to be the center of the whole image, and people are usually more interested in the face area in the image according to the visual characteristics of the human eye. When a person is identified, the scene is first retrieved to determine the location of the face, in which motion and color are the main clues. And then the area of interest is detected for the face, which is the task of the face detection. Then, the face is followed. This is the task of face tracking[1]. It is not until a recognition-friendly pose appears that it will determine whether the face is familiar. This process is very simple for people, but it has always been a problem in the field of computer vision. To solve the problem of face detection, there are the following difficulties: (1) the face is non-rigid, with different expressions. Facial expressions will lead to changes in the mouth, eyes and other positions and shapes. (2) The face will present different gestures, the face front, side, rotation, etc. The difference between elevation angles is very large, and the eyes, nose and mouth may be partially or completely occluded; (3) Some special structures are not every face, such as glasses, hair, beard, etc.; (4) The difference of facial features between different sex, age and race is very big; (5) The influence of light intensity and angle on the change of light condition will result in the problem of uneven brightness, shadow and so on; (6) The influence of complex background. There may be areas in the background of complex images that may be misjudged as faces; (7) Effects of camera imaging. In the process of shooting, the performance of the camera, such as pixel, sensitivity, focus, will affect the face in the image rendering. The technical difficulty in face detection is also the technical difficulty of face tracking. In addition, due to the need of practical application, face tracking algorithm requires good robustness and real-time performance. Robustness means that the algorithm can track the moving face continuously and stably in all kinds of environments.

The main reasons that affect the robustness of the tracking algorithm are[2] the attitude change of the tracked face, the illumination change of the environment, the irregular deformation of the target caused by partial occlusion and the temporary disappearance of the target caused by total occlusion. Real-time is to point to a practical face tracking system must be able to achieve real-time tracking of moving faces.

However, the object of visual tracking algorithm is video containing huge amount of data. These algorithms often require a lot of computing time and are difficult to meet the real-time processing requirements. Generally, the simple algorithm can realize real-time tracking, but the tracking precision is very low. The complex algorithm has high tracking precision, but the real-time is very poor. To sum up, face detection and tracking technology is a challenging research direction. At present, all theories and algorithms approach human's recognition ability from all aspects.

DOI: 10.25236/icmcs.2019.050

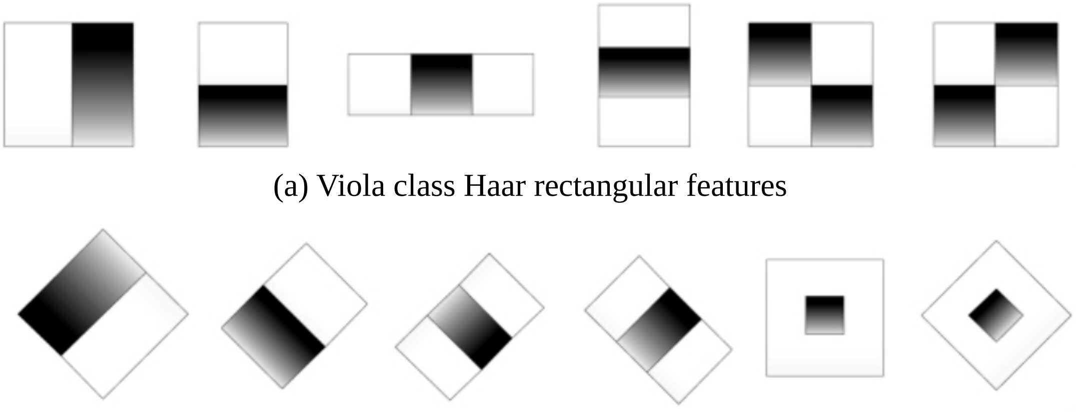
Application

Compared with other biometric recognition techniques, face features are more prominent, often without the active cooperation of recognition objects, which can be collected without their knowledge. Other biometric identification methods, such as fingerprint recognition, palm recognition and eye iris recognition, require active coordination of recognition objects, which are difficult to obtain in many environments and situations. In addition, human face is the most important part of interactive communication and other routine activities. Therefore, face detection and tracking technology has a wide application prospect. Face detection and tracking can be used in intelligent video surveillance system. The most common is surveillance of homes, parking lots, public places, banks, etc. to prevent theft and sabotage.

In the field of human-computer interaction, future machines are expected to communicate easily with humans, and machines must have the ability to perceive the state of human emotion if they are to respond correctly to human instructions. Machines shall have the capacity to sense the human feeling status[3]. The research in document shows that the most important way to obtain human emotion information is through the facial expression state in video stream. Face detection and tracking is the key step to accomplish this task. It is the basis of further combining facial expressions to study people's recognition and understanding of behavior, such as gesture-driven control, sign language translation, body posture and gesture[4-6]. Most digital cameras today use face detection technology to achieve automatic focusing and exposure. Many of the best photo management software include face detectors, such as Apple's iPhoto, Google's Picasa and the soft Windows Live Photo Gallery, which help better mark and organize photos.

Face Detection

Face detection is a process in which a given image or a set of image sequences are searched with a certain strategy to determine the location and region of all faces, and determine the number of faces and spatial distribution process[7]. In the past decade, great progress has been made in the field of face detection, especially since Viola algorithm[8] was proposed. There are three core ideas in Viola algorithm: (1) Using cascaded method, a single classifier can be synthesized into cascaded classifier, so that the non-face area can be quickly discarded, and more attention is paid to the calculation of the possible face area; (2) The concept of "integral diagram" is proposed, which can quickly calculate the features used by the classifier; (3) The most critical features selected from a large number of feature sets are selected by the training algorithm based on AdaBoost (Adaptive Boosting). Generally speaking, the appearance-based approach first collects a large number of face and non-face samples, then uses a machine learning algorithm to learn a face model, and builds a classifier to retrieve faces in the image. In this process, there are two important steps, one is to select which features. The other is to apply what sort algorithm. Many scholars have improved the original Viola algorithm in terms of feature selection and classification algorithm respectively, which can improve the accuracy of face detection as well as the speed of detection.



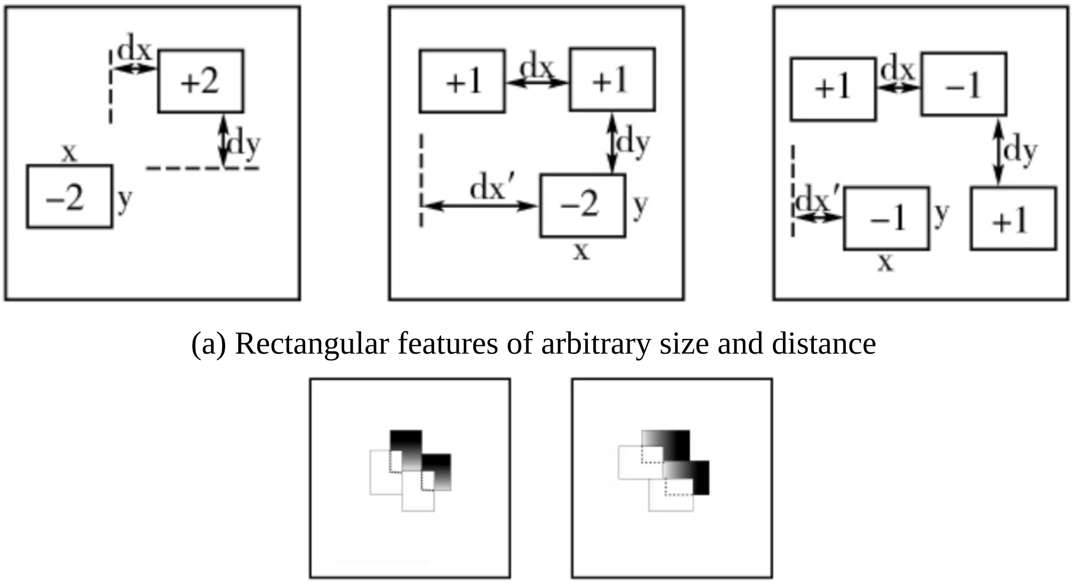
(a) Viola class Haar rectangular features

(b) Deformed Haar-like Rectangular Feature

Figure. 1 Haar Rectangular Feature

The Haar-like feature used in Viola algorithm is a rectangular feature. As shown in Figure 1 (a), the eigenvalue of Haar-like feature is quickly calculated by using integral graph. Because these features are very effective in front face detector, many scholars have been interested in extending the original Haar-like features by changing the combination of rectangular features. Lienhart and Maydt[9] extend the Haar-like feature, introducing the feature of rotation 45 degrees angle and the feature of center encircling, as shown in Figure 1 (b), so that it can be used for face detection with slightly tilted plane. At the same time, the integral graph calculation method of this new extended feature is given.

Given the limitations of Haar-like features in multi-view face detection, as shown in Figure 1 (a), in the document[10], the author gives a greater degree of freedom to the rectangular region. The three features shown in Figure 2 (a) show that the rectangle is x × y in size. The distance between them is defined as (dx, dy), and these are asymmetric features that detect asymmetric features in side faces. Jones and Viola also try to improve the Haar-like feature. They propose a diagonal filtering feature in literature[11], as shown in Figure 2 (b). The eigenvalues of diagonal filtering can also be calculated by multiple query integral graphs. All of these improvements are based on static images, which raise the challenge of detecting faces in video, but also provide some useful motion information. Jones et al.[12] extended the Haar-like feature to be used in video human detection. The author defines five motion filters, and the Haar feature acts on the filtered image of the motion filter. In this way, the motion information and the appearance information in the video are combined, and the same as the Viola algorithm, the AdaBoost framework training detector is used. The result is a very low FP (FALSE POSITIVE) rate.



(a) Rectangular features of arbitrary size and distance

(b) Diagon Filtering Rectangular Feature Diagram 2

Figure. 2 The rectangle features

The joint Haar feature is proposed by Mita et al.[13] The so-called Joint Haar-like features actually refer to the coexistence of multiple Haar features. The joint Haar feature uses a similar characteristic computation and threshold model to the original Haar feature. But the output of the joint Haar feature is a binary combination with 2F possible, in which F is the number of features of the combination. As shown in Figure 2, j is the decimal number corresponding to the combination of F binary features. Compared with the original Viola detector, which uses single-class Haar features to train each weak classifier, combined Haar features can better capture face features, and a more powerful classifier can be established.

In addition, Levi et al. proposed the feature of local edge histogram, which will calculate the edge direction histogram in the sub-region of the test window. And it is more powerful than Haar-like features in capturing the geometric attributes of the face. Baluja et al proposed the use of simple pixels on features, Abramson et al proposed the use of a set of control points set of relative value features. These two pixel-based features is much less than the original type of Haar in calculation volume, so the computation speed has greatly improved. But their ability to distinguish for the establishment of a high-performance detector is not enough. At the same time, some scholars use some shape features to model face targets.

Face Tracking

Face tracking is usually based on face detection. In the video tracking face, the first step is to detect the face target. Plus, the face may be occluded or removed from the video, when it also needs face detection to re-locate the face. So, the two are inseparable.

From the point of view of common face tracking technology, face tracking can usually be divided into: Model-based approach, motion information-based approach, face local feature-based approach and neural network-based approach. Literature[14], starting from these four categories, made a more detailed on the current face tracking technology for each specific method description and comparison, therefore this paper no longer repeated. The following is to solve the problem of long-term face tracking in any environment to gradually analyze the research situation of face tracking.

From the point of view of common face tracking technology, face tracking can usually be divided into: Model-based approach, motion information-based approach, face local feature-based approach and neural network-based approach. Literature[14], from these four categories, made a more detailed description and comparison on the current face tracking technology for each specific method, therefore this paper no longer repeated. The following is to solve the problem of long-term and long-term face tracking in any environment to gradually analyze the research situation of face tracking. Long-term face tracking in any environment is a challenging problem. For example, for a particular face, there may be frame loss, sudden target deformation, target being blocked for a long time and other complex situations. There are two kinds of modeling methods for tracking target faces: Static and dynamic methods. The static method[15] assumes that changes in the face's appearance are limited or known, so when an unexpected deformation of the face occurs, the trace fails. The dynamic method[16] solves this problem by updating the face target model in real time during the process of tracking, but there is a potential assumption in it: Each update is correct. Each error update will cause errors in the model, and over time it will result in tracking drift. The problem of face tracking has been solved by using Visual Constraints (Visual Constraints). Although this method has been proven to increase robustness and accuracy, its performance is only tested in videos where the face is not out of sight. If someone's face is moved in and out of the video, it is necessary to re-locate the face. Otherwise it can't achieve long-term face tracking. In recent years, the tracking field has attracted a lot of attention, such as the 2009 year Kalal and so on[15-16] proposed on-line target tracking framework TLD (Tracking-Learning-Detecting). The TLD framework decomposes long-term tasks that track unknown targets into tracking, learning, and detection problems. Firstly, it locates the target in the first frame, and then determining the location, scope, or absence of the target in each subsequent frame. The tracker uses the motion information of the target in the video to track the frame to the frame's target; The detector locates all observed external features and corrects the tracker when necessary; and learns to evaluate the detector's error and correct it to avoid the same error again. The TLD system is a semi-supervised on-line system, and the detector does not require off-line training. Then Kalal built a face tracking system based on TLD. This system extends TLD with the concept of verifier and general detector to realize real-time face tracking. The detector is an offline-trained Viola face detector used to locate faces. The verifier is trained online to determine which faces match the target being tracked.

The experiment proves that the system can continuously track the face to 35471 frame in the video. The goal of long-term face tracking in any environment is basically realized. However, the face tracker is only a single face tracker, and the real-time tracking of multi-faces based on TLD framework remains to be further studied. At present, the research on face tracking system has gradually entered the application stage, and the application field of face tracking is more and more extensive.

Conclusion

This paper introduces the research status of face detection and face tracking, mainly focus on the key stages, important methods and recent research hot spots of face detection and face tracking. Face detection and tracking technology is an important and difficult subject in the field of computer vision. Although a breakthrough has been made in recent years, there is still much room for improvement in the accuracy and speed of face detection and tracking. The author holds that using external information is a way to improve the performance of face detection and tracking, as the human face is closely related to other parts of the body. These parts can provide a clue to locate the face, which deserves further research and exploration. It is believed that there will be a better face detection and face tracking system in the near future.

References

[1] Multi-channel SSVEP Pattern Recognition Based on MUSIC[A]. Kun Chen, Quan Liu, Qingsong Ai.Selected, peer reviewed papers from the 4th International Conference on Intelligent Structure and Vibration Control(ISVC 2014)[C]. 2014.

[2] Hager G D, Belhumeur P N. Efficient region tracking with parametric models of geometry and illumination[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(10) : 1025-1039.

[3] Sun Y, Sebe N, Lew M, et al. Authentic emotion detection in real-time video[C] / /Proceedings of ECCV Workshop on HCI. 2004: 94-104.

[4] Pavlovic V I, Sharma R, Huang T S. Visual interpretation of hand gestures for human-computer interaction: A review[J] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997,19(7) : 677-695.

[5] Lien C C, Huang C L. Model-based articulated hand motion tracking for gesture recognition[J] . Image and Vision Computing, 1998, 16(2) : 121-134.

[6] Choi H I, Rhee P K. Head gesture recognition using HMMs[J] . Expert Systems with Applications, 1999, 17 (3) :213-221.

[7] Yang M H, Kriegman D J, Ahuja N. Detecting faces in images: A survey[J] . IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1) : 34-58.

[8] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C] / /Proceedings of the 2001 IEEE Computer Society Conference on CVPR. 2001: 511-518.

[9] Lienhart R, Maydt J. An extended set of Haar-like features for rapid object detection[C] / /Proceedings of 2002 International Conference on Image Processing. 2002: 900-903.

[10] S. Li, L. Zhu and Z.Q. Zhan, et al. Statistical learning of multi-view face detection[C] / /Proceedings of the 7th European Conference on Computer Vision. 2002: 67-81.

[11] Jones M, Viola P. Fast Multi-view Face Detection[R]. Mitsubishi Electric Research Lab TR-20003-96, 2003.

[12] Jones M, Viola P, Jones M J, et al. Detecting pedestrians using patterns of motion and appearance[J]. InternationalJournal of Computer Vision, 2005, 63(2) : 153-161.

[13] Mita T, Kaneko T, Hori O. Joint Haar-like features forface detection[C] / /The Tenth IEEE International Conference on Computer Vision. 2005: 1619-1626.

[14] Z.Y. Li, Q. liu and X.M. Liu. Review of facial tracking methods[J]. Journal of Yanan University (Natu ral Science Edition), 2005,24(4):39-44.

[15] Kalal Z, Matas J, Mikolajczyk K. Online learning of robust object detectors during unstable tracking[C] / /Proceedings of the IEEE On-line Learning for Computer Vision Workshop. 2009: 1417-1424.

[16] Kalal Z, Matas J, Mikolajczyk K. PN learning: Bootstrapping binary classifiers by structural constraints[C] / /Proceedings of the Conference on Computer Vision and Recognition. 2010: 49-56.