Automatic Human Face Detection and Tracking

AMOL DABHOLKAR and KIN CHOONG YOW
School of Computer Engineering, Nanyang Technological University,
Nanyang Avenue, Singapore 639798

Abstract: - With the growing interest in human-computer interaction, teleconferencing and visual surveillance, many automated human face detection and tracking system have been developed. In this project, a human face detection system is combined with tracking algorithms to form an automated human face detection and tracking system. The detection module makes use of skin color classification and segmentation to estimate the boundary of the face. Each pixel is classified as “skin” or “non-skin” colour and the largest cluster of “skin” color is labeled as a possible face candidate. In the tracking module, facial features are tracked using motion-based and model-based algorithms. A Kalman filter is employed to estimate the feature motion. Various tests are conducted and the system is shown to perform reasonably well for images in an upright and frontal view. For other orientations and views of the face, the system may lose track of some of the features. The integrated system acts as the first stage in finding the face and tracking its features’ position. Other applications can then make use of this information for human-computer interaction.

Key-Words: - Computer Vision, Human Face Detection, Tracking, Kalman Filters

1 Introduction

The invention of computers and the development of artificial intelligence systems have made the study on human body parts easier and faster. Many researchers are able to model the functions and capabilities of human body parts with computer, sensors and actuators. Different types of sensors are used to simulate the various sensors of the human body parts, e.g., cameras as eyes and microphones as ears. Motors and actuators have also been invented to replace human muscles for performing a variety of jobs.

Computers have been widely used in research involving the human visual system, such as computer vision and pattern recognition. One area of interest is human face detection, recognition and tracking. The human face is one of the most prominent and unique feature of human beings. It plays a significant role in the identification of people, and in the perception of emotional expression. Among the many methods of identification (voice, thumb print, DNA, etc.), the face is one of the easiest means of identification. This is because it is easily accessible and provides a good degree of accuracy.

The aim of this work is to use a computer to simulate the human visual system to detect and track human faces on digital images. Automated face detection system is the fundamental module in many artificial systems. Human Computer Interface (HCI), video conferencing, security identification/ verification and surveillance systems are some common applications that can make use of this automated face detection and tracking system.

2 Related Work

Numerous research on human face detection and tracking has been carried out using different approaches. However, the problem of face detection and tracking is so complex that there is no single method that would guarantee reliable results under all circumstances. We will now look at the various approaches that have been undertaken by researchers to tackle the face detection and the face tracking problem.

2.1 Colour-based Approaches

The colour of human skin can be used to provide a good initial estimate on the location of the face. With this information, other facial features can then be located. This approach labels each pixel according to its similarity to skin colour. Regions of skin colour pixels are then tested to satisfy the symmetry of a face [3, 4]. The problem with this method is that the background or other exposed body parts may also be selected. Algorithms based on skin colour would fail since the skin colour region cannot be obtained reliably by image segmentation [3].

2.2 Feature-based Approaches

The feature-based algorithm operates by first detecting the facial features and studying their mutual distances using statistical model. Next, a set of facial features is located using local feature detectors. These features are grouped using a probabilistic model into face candidates depending on their geometrical relationship [10, 14]. This approach is unable to work under different imaging conditions because the image structure varies too much to be robustly detected. Moreover, if the algorithm fails to detect one of the crucial facial features, it will be unable to find symmetrical relationship among the features and hence fails to detect a face [14].

2.3 Fuzzy Logic Approaches

Face-like regions are extracted by a fuzzy pattern matching approach using predefined face models [3]. However, research in this field is still in its nascent fuzzy stages. Fuzzy logic might give desirable results for the problem of pattern recognition as the matching is more flexible, but the process of arriving at the solution is still irrational.

2.4 Neural Network-based Approaches

Neural networks are information processing structures that try to simulate the functioning of the human brain and are “trained” using sample scenes rather than programmed in a classical way. In recent works, the face is detected by sub-sampling different regions of the image to a standard sized image and passed through a neural network filter [11, 8]. The algorithm performed well for frontal view of faces but is difficult to extend to different views of the face. If the network is trained adequately, the performance of the detection can be quite high. However, a neural network method of processing data in response to a stimulus cannot be guaranteed to be reliable, as it is not based on logic but on how well it learns from the training samples.

2.5 Template-based Approaches

The template based approach uses different face models at different coarse-to-fine resolutions. The image is searched at the coarsest scale first and once a match is found, the image is searched at the next finer scale until the finest scale is reached [12, 7]. Only one model is used at each scale, usually in fronto-parallel view and is difficult to extend to other views. Models of explicit shapes only work well in high resolution and relatively noise free images [15].

2.6 Alpha-beta Filter

The alpha-beta filter enables the prediction of the next state in a state-transition system to be adjusted according to the error between the current estimation and measurement. However, the percentage of the error (between measured and expected used to adjust the model parameters) is kept constant for the entire image sequence or is adjusted by the programmer.

2.7 Kalman Filter

Another popular filter is the Kalman filter [1], which differs from other statistical filters in that it enables the weight of the error influence to be adjusted automatically as the confidence in the parameters and model of behaviour increased. The predicted value and the measured value are used to adjust the filter’s parameters. However, the filter can only optimally weight the predictions and the measurements for zero-mean Gaussian noise and process noise encountered in visual tracking.

2.8 Deformable Template Matching

The deformable template matching uses the model based approach. Multiple templates are used for a single feature, with each template corresponding to a different view of the feature. This provides a high degree of accuracy but the tradeoff is that it is computationally expensive.

3 Human Face Detection

Our system consists of two modules: a face detection module and a face tracking module. The objective of the face detection module is to locate the face boundary and extract the facial features. The positions of facial features are required in the tracking module. Thus, face detection is considered as the initialization stage for the system. Face detection is performed by two separate approaches, a colour –based approach followed by a feature-based approach.

A skin colour classification algorithm is used to classify the pixels in the image as either “skin” or “non-skin”. Many researchers have used skin colour as an important cue for detecting faces. They reported that a skin colour model is reasonably stable and can be used to detect faces with varying skin colours [13, 4, 3, 2]. The skin colour model can either be a fixed set or created using portions of the face. Although a fixed skin model is sufficient, the colour distribution will be quite large. This will result in more pixels being labelled as skin colour.

If the image sequence is constrained by a fixed illuminating source, a more specific skin model created for that particular set of images would give better results. Figure 1 shows the results of using a general skin colour model and figure 2 shows the difference that can be obtained if a specific skin colour model (obtained from a portion of the face) is used.

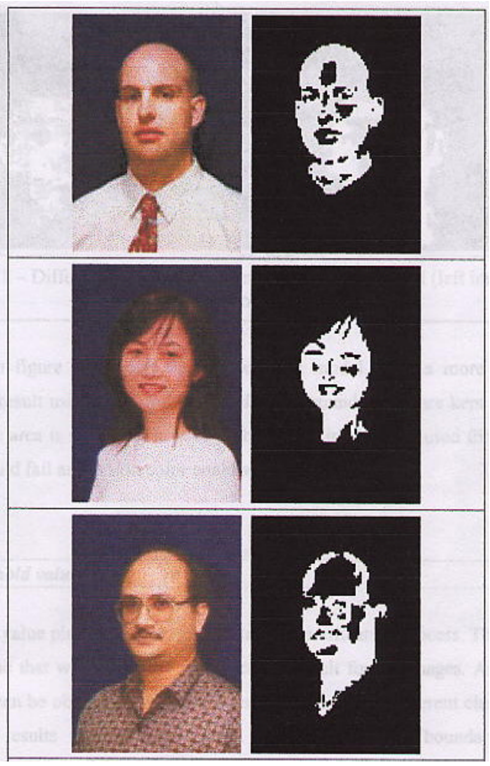


Figure 1. Results Of Skin Colour Classification Using A General Skin Colour Model

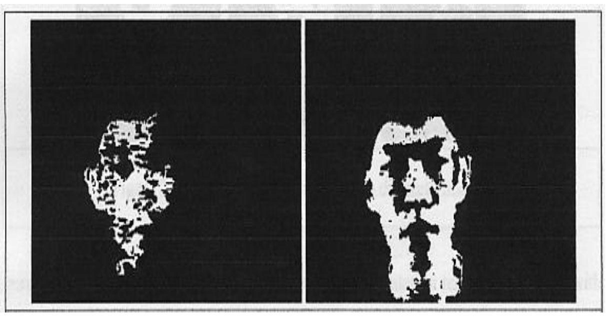


Figure 2. Different Results Using A General Skin Colour Model (Left Image) And A Specific Skin Colour Model (Right Image).

The "skin" regions are then clustered using region growing techniques [6] and a most probable face candidate is determined from the largest skin cluster. After locating a most probable face candidate, the holes in the segmented image are analyzed to give an approximate position of the facial features (e.g. eyes and mouth). Facial features are usually classified as non skin-colour regions and can be analyzed based on its size, orientation and distance from the boundary [9]. Gaussian derivative filters [14] are subsequently used to extract the exact coordinates of the facial features based on the approximate position.

After finding the coordinates of the facial features (eyes, nose and mouth), a confirmation check is conducted based on the symmetrical relationship of these features. Some tolerance is allowed as the person's face may be facing at an angle. For example, if the system is unable to find the eyes initially, it tries to recover them using information on the locality of the nose and mouth. A search for the darkest pixel is conducted using a calculated coordinate: a position directly above the edge of the mouth, of distance twice the difference from nose to mouth. Figures 3 and 4 show the calculated search area for missing features (eyes) and the location of the feature after the correction process.

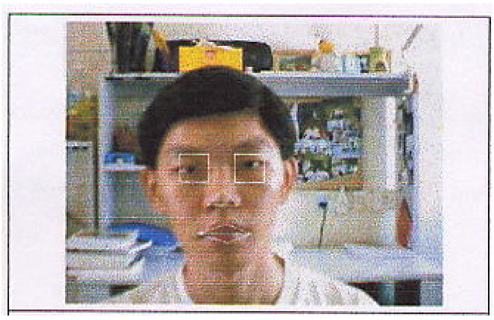


Figure 3. Calculated Search Area For Missing Eyes Recovery (The 2 Boxes)

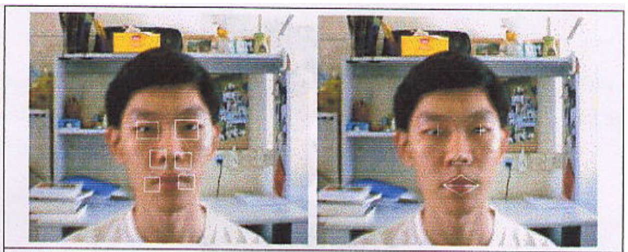
Figure 4. Location Of Facial Features After The Correction Process

4 Human Face Tracking

The second module in the system is the face tracking module. As the tracking is required to operate in real-time, a fast tracking algorithm is required. Kalman filter is selected as it guarantees an optimal estimate of the parameter, such as the position of a feature in the next frame of a sequence. It has the ability to adjust its own parameters according to the statistics of the measurements from each image. All the parameters of the Kalman filter have covariance, which indicates the reliability of the data. The output also has a covariance so that the tracking system does not only obtain an estimate but also the reliability of the estimate.

In our proposed tracking system, Kalman filters are applied to the six facial features' positions (namely, the centre of both eyes, the corners of the nose and the corners of the mouth) detected by the face detection module. The filter will estimate the likely positions of these features in the current frame based on its position over the last two frames. In this way, a small search window can be generated to locate the actual position of the feature.

Figure 5 shows the search areas for feature extraction and the results of the Kalman filter. The second row of figure 5 shows the predicted, measured and final values of the Kalman filter. The predicted position can be different from the measured position, but due to the ability to minimize error covariance, the final output of the Kalman filter is close to the actual measured position. The error in the prediction in the next frame will also be minimized.



Feature	X-coordinate Predict	Measured	Final	Y-coordinate Predict	Measured	Final
L eye	141.00	135.00	137.50	117.87	120.00	119.11
R eye	191.54	183.00	186.55	118.91	118.00	118.38
L nose	158.63	149.00	153.00	156.86	153.00	154.61
R nose	181.86	172.00	176.10	155.95	151.00	153.05
L mouth	147.88	141.00	143.87	183.13	177.00	179.56
R mouth	189.68	183.00	185.77	172.96	174.00	173.56

Figure 5. Left image shows boundary of search window for the features; right image shows the final result of Kalman filter.

The use of Kalman filter produces a fast tracking system. However, it is unable to determine the correctness of the object that it is keeping track. To overcome this, Random Sample Consensus (RANSAC) [5] is incorporated into the tracking system so that there is some form of error checking capability. The idea behind the RANSAC technique is to find a significant group of points that are consistent with a particular pose and reject the remaining points as outliers.

In the tracking module, the distance between each feature is used to determine if any feature is out of position. Since there is a total of six feature points, a total of 15 distances are measured. Five distances (consensus sets) are measured from each feature and if more than half (three and above) of the distances are not within the accepted range, the feature points is labelled as an outlier and corrective action is taken.

As the tracking progress, features could be detected out of position. If there is no corrective action taken, the point may never be recovered. The incorporation of RANSAC helps to recover outlier points. When the point gets occluded, the algorithm is able to keep track of its expected position so that it can be recovered later.

Figures 6, 7 and 8 shows the result of the face tracking module after incorporating the RANSAC algorithm into the Kalman filter.

Figure 6. Output of a horizontal face movement

Figure 7. Output of a face being tilted

Figure 8. Output of a face being partially occluded.

5 Conclusion and Future Work

We have developed a human face detection and tracking system that comprises of four main approaches: colour classification and segmentation, facial features extraction, Kalman filter prediction and random sample consensus error checking. These different approaches are combined together to forma a fast, simple yet reliable face detection and tracking system.

The performance of the integrated system is analyzed by inputting different test sequences of images. From the test results, we can conclude that the system is able to detect faces of near frontal view, with the facial features clearly visible in the image. The system is also able to track the facial features reliably if the face is within the image, even if a feature has been occluded. However, the tracking system may fail if there is a sudden, large, movement within a single frame sequence.

Future work will be directed towards using the ambient (or background) information to perform image subtraction. This will produce an image difference which will show the outline of the person in the image very clearly. Using this outline together with the colour and feature extraction modules, we can reliably detect a person in the image.

The system should also be able to automatically switch between the detection and tracking modes in a sequence. If there is no person in front of the camera, the system should be running in detection mode. If a person enters the scene and is successfully detected by the algorithm, the system should automatically switch to tracking mode.

References:

[1] S. M. Bozic. *Digital and Kalman Filtering*. Edward Arnold Publishers, 1986.

[2] T. Chang, T. Huang, and C.Novak. "Facial Feature Extraction from Colour Images". *In Proceedings of the International Conference on Pattern Recognition*, pp. 39-43, 1994.

[3] Q. Chen, H. Wu and M. Yachida. "Face Detection by Fuzzy Pattern Matching". In Proceeedings of the IEEE International Conference on Computer Vision, Piscataway, USA, pp. 591-596, 1995.

[4] Y. Dai and Y. Nakano. "Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene". *Pattern Recognition*, vol. 29(6), pp. 1007-1017, 1996.

[5] M. A. Fischler and R. C. Bolles. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography". *Communications of the ACM*, vol. 24(6), pp. 381-395, 1981.

[6] R. Jain, R. Kasturi and B. G. Schunck. *Machine Vision*, McGraw Hill, Singapore, 1995.

[7] A. Lanitis, C. J. Taylor and T. F. Cootes. "An Automated Face Identification System using Flexible Appearance Models". *Image and Vision Computing*, vol. 13(5), pp. 393-401, 1995.

[8] H.A. Rowley, S. Baluja and T. Kanade. *Human Face Detection in Visual Scenes*. Technical Report CMU-CS-95-158R, Carnegie Mellon University, Nov. 1995.

[9] E. Saber and A. M. Tekalp. "Frontal-view Face Detection and Feature Extraction using Color, Shape and Symmetry Based Cost Functions". *Pattern Recognition Letters*, 19, pp. 669-680, 1998.

[10] Y. Sumi and Y. Ohta. "Detection of Face Orientation and Facial Components using Distributed Appearance Modelling". *In Proceedings of the International Workshop on Automatic Face and Gesture Recognition*, pp. 254-259, Zurich, 1995.

[11] K. K. Sung and T. Poggio. "Example-Based Learning for View-Based Human Face Detection". *IEEE Transactions on Pattern*

Analysis and Machine Intelligence, vol. 20(1), pp. 39-50, January 1998.

[12] G. Yang and T. Huang. "Human Face Detection in a Scene". *IEEE Transactions*, pp. 453-457, 1993.

[13] J. Yang and A. Waibel. "A Real Time Face Tracking System". Technical Report Computer Science Department, Carnegie Mellon University, 1997.

[14] K. C. Yow and R. Cipolla. "A Probabilistic Framework for Perceptual Grouping of Features for Human Face Detection". *In Proceedings 2nd International Conference on Automatic Face and Gesture Recognition*, pp 16-21, Vermont, 1996.

[15] A.Yuille, P. Hallinan and D. Cohen. "Feature Extraction using Deformable Templates". *International Journal of Computer Vision*, vol. 8(2), pp. 99-111, 1992.