Automatic Human Face Counting in Digital Color Images

MONA F.M. MURSI [1,2]
monmursi@coeia.edu.sa

GHAZY M.R. ASSASSA [1,3]
gassassa@coeia.edu.sa

ABEER AL-HUMAIMEEDY [1,2]
humaimeedy@coeia.edu.sa

KHALED ALGHATHBAR [1,4]
ghathbar@coeia.edu.sa

[1] Center of Excellence in Information Assurance (CoEIA),
[2] Department of Information Technology
[3] Department of Computer Science
[4] Department of Information Systems
College of Computer and Information Sciences
King Saud University
Kingdom of Saudi Arabia

Abstract: - Automatic human face detection is considered as the initial process of any fully automatic system that analyzes the information contained in human faces (e.g., identity, gender, expression, age, race and pose). In this paper, color segmentation is used as a first step in the human face detection process followed by grouping likely face regions into clusters of connected pixels. Median filtering is then performed to eliminate the small clusters and the resulting blobs are matched against a face pattern (ellipse) subjected to constraints for rejecting non-face blobs. The system was implemented and validated for images with different formats, sizes, number of people, and complexity of the image background.

Key-words: - Face detection, color segmentation, face pattern (ellipse).

1 Introduction

Face detection deals with finding faces in a given image and hence returns the face location and content. The problem of automatic human face detection can be stated as follows: Given a still or video image, detect and localize an unknown number of faces, if any. The problem of human face detection is a complex and highly challenging one having a variety of parameters including face color, scale, location, size, pose (frontal, profile), orientation (up-right, rotated), illumination, partial occlusion, cluttered scenes, image background and camera distance. Face detection is one of the visual tasks which humans can do effortlessly. However, in computer vision terms, this task is not trivial. Face detection has various applications in many areas, an important example of which is security related surveillance in confined areas. Once an image is analyzed for faces detection, the faces in the image are tallied. There are many related research problems of face detection; the list includes: (1) face localization: determination of the image position of a single face [1], (2) facial feature detection: detecting the presence and location of features such as eyes, nose, nostrils, eyebrow, mouth, lips, ears, etc [2, 3], (3) face recognition: compares an input image against a database (gallery) and reports a match, if any, (4) face tracking: continuously estimate the location and possibly the orientation of a face in an image sequence in real time , and (5) facial expression recognition: identifying the affective states , e.g., happy, sad, wondering, frightened, disgusted. The solution to the face detection problem involves various areas in image processing including segmentation, extraction, and verification of faces and possibly facial features for an uncontrolled background. Segmentation schemes have been presented, particularly those using motion, color, and generalized information. A survey on pixel-based skin color detection techniques was presented in [4]. The use of statistics and neural networks has also enabled faces to be detected from cluttered scenes at different distances from the camera [5. Many algorithms have been implemented for facial detection, including the use of color information, template matching, neural network, edge detection and Hough transforms [5, 6].

This paper is organized as follows: section 2 presents related work; section 3 discusses the proposed face detection approach; in section 4, test results are presented and discussed. Finally, section 5 concludes this paper.

2 Related Work

Face detection (FD) methods require a priori information of the face; these methods can be generally classified under two broad categories, with some methods overlapping the boundaries of these categories: feature-based and appearance-based methods.

2.1 Feature-based Approach

This approach makes explicit use of face knowledge and follows the classical detection methodology in which low level features are derived prior to knowledge-based analysis [7]. The apparent properties of the face such as skin color and face geometry are exploited at different system levels. Typically, in these techniques face detection tasks are accomplished by manipulating distance, angles, and area measurements of the visual features derived from the scene. Since features are the main ingredients, these techniques are termed the feature-based approach.

2.1.1 Low-level Analysis

This heading includes edges, gray information, color, and motion as briefly introduced hereafter.

Edges: Edge detection is the foremost step in deriving edge representation [8]. So far, many different types of edge operators have been applied; the Sobel operator being the most common filter. In an edge-detection-based approach to face detection, edges need to be labeled and matched to a face model in order to verify correct detections [9].

Gray Information: Besides edge details, the gray information within a face can also be used as features. Facial features such as eyebrows, pupils, and lips appear generally darker than their surrounding facial regions [6].

Color: It was found that different human skin color gives rise to a tight cluster in color spaces even when faces of difference races are considered [10]. This means color composition of human skin differs little across individuals. Since the main variation in skin appearance is largely due to luminance change (brightness), normalized RGB colors are generally preferred, so that the effect of luminance can be filtered out.. In skin color analysis, a color histogram based on r and g shows that face color occupies a small cluster in the histogram. By comparing color information of a pixel with respect to the r and g values of the face cluster, the likelihood of the pixel belonging to a flesh region of the face can be deduced.
HSI color representation has advantages over other models in giving large variance among facial feature color clusters[11]; it is used to extract facial features such as lips, eyes, and eyebrows.
Color segmentation can basically be performed using appropriate skin color thresholds where skin color is modeled through histograms charts [12, 13]. More complex methods make use of statistical measures that model face variation within a wide user spectrum [10].

Motion: A straightforward way to achieve motion segmentation is by frame difference analysis. This approach, whilst simple, is able to discern a moving foreground efficiently regardless of the background content [6].

2.1.2 Feature Analysis

Features generated from low-level analysis are likely to be ambiguous. For instance, in locating facial regions using a skin color model, background objects of similar color can also be detected. This is a classical many-to-one mapping problem which can be solved by higher level feature analysis. In many face detection techniques, the knowledge of face geometry has been employed to characterize and subsequently verify various features from their ambiguous state. There are two approaches in the application of face geometry: (1) the first involves sequential feature searching strategies based on the relative positioning of individual facial features, (2) grouping features as flexible constellations using various face models.

Feature searching: The detection of the prominent features then allows for the existence of other less prominent features to be hypothesized using anthropometric measurements of face geometry. For instance, a small area on top of a larger area in a head and shoulder sequence implies a "face on top of shoulder" scenario, and a pair of dark regions found in the face area increase the confidence of a face existence. Among the literature survey, a pair of eyes is the most commonly applied reference feature [6] due to its distinct side-by-side appearance.

Constellation analysis: Some of the algorithms mentioned in the last section rely extensively on heuristic information taken from various face images modeled under fixed conditions. If given a more general task such as locating the face(s) of various poses in complex backgrounds, many such algorithms will fail because of their rigid nature. Later efforts in face detection research address this problem by grouping facial features in face-like constellations using more robust modeling methods such as statistical analysis[14].

2.2 Image-based Approach

Face detection by explicit modeling of facial features, presented above, is troubled by the variation of face appearance and environmental conditions. There is still a need for techniques that can perform in more hostile scenarios such as detecting multiple faces with clutter-intensive backgrounds. This requirement has inspired a new research area in which face detection is treated as a pattern recognition problem. The basic approach in recognizing face patterns is via a training procedure which classifies examples into face and non-face prototype classes. Comparison between these classes and a 2D intensity array (hence the name image-based) extracted from an input image allows the decision of face existence to be made. The simplest image-based approaches rely on template matching [15], but these approaches do not perform as well as the more complex techniques presented in the following sections. Most of the image-based approaches apply a window scanning technique for detecting faces. The window scanning algorithm is in essence just an exhaustive search of the input image for possible face locations at all scales.

The image-based approaches may be classified into linear subspace, neural networks, and statistical methods as briefly described below.

Linear Subspace Methods- PCA: Images of human faces lie in a subspace of the overall image space. To represent this subspace, many methods may be applied such as the principal component analysis (PCA) [16, 17]. Given an ensemble of different face images, the technique first finds the principal components of the distribution of faces, expressed in terms of eigenvectors of the covariance matrix of the distribution. Each individual face in the face set can then be approximated by a linear combination of the largest eigenvectors, more commonly referred to as eigenfaces, using appropriate weights

Neural Networks: Neural networks have become a popular technique for pattern recognition problems, including face detection. The first neural approaches to face detection were based on Multi Layer Perceptrons (MLPs) [5]. Ref [18] used possible face features (eyes, and mouth) and pass them to a neural network to confirm face validation.

Statistical Approaches: There are several statistical approaches to face detection including support vector machine (SVM) and Bayes' decision rule. In [19], a system was developed for real-time tracking and analysis of faces by applying the SVM algorithm on segmented skin regions in the input images to avoid exhaustive scanning. SVMs have also been used for multi-view face detection by constructing separate SVMs for different parts of the view sphere [20]. In [21] Rong Xiao] SVM-filter and color-filter are applied to refine the final prediction of face detection.

Recently in [22] Farajzadeh, Nacer 2008], a robust face and non-face discriminability was achieved via the introduction of a hybrid system for face detection based on Viola and Jones's work and the usage of Radial Basis Neural Network.

ISSN: 1790-5117

ISBN: 978-960-474-054-3

3 The Proposed Face Detection Approach

This section describes the main steps used in the proposed face detection method; this includes noise reduction, skin color segmentation, and morphological processing. Some details are given below, more details may be found in [23].

3.1 Noise Reduction

The goal of noise reduction is to reduce the number of non-facial blobs in the image. Two algorithms were considered for noise removal, (1) the median filter as applied in the spatial domain of the image, and (2) the band-pass filter as applied in the frequency domain of the image. Numerical Experiments showed that the two algorithms yielded approximately similar results. The median filter was selected in the present work because it showed faster performance.

3.2 Skin Color Segmentation

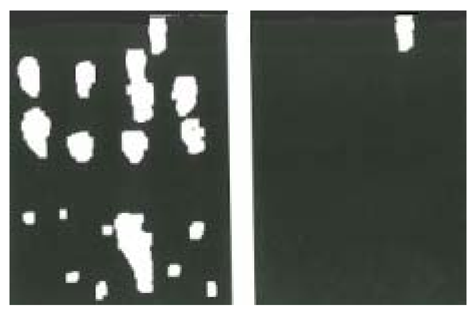
With the assumption of a typical photographic scenario, it would be clearly wise to take advantage of face-color correlations to limit our face search to areas of an input image that have at least the correct color components. We used the hue-saturation-value (HSV) color space because it is compatible with the human color perception. Hue (H) is represented as an angle, the purity of colors is defined by the saturation (S), which varies from 0 to 1. The darkness of a color is specified by the value component (V), which varies also from 0 (root level) to 1 (top level). Hue and saturation were taken as discriminating color information for the segmentation of skin-like regions with the following values [24]: S(min) = 0.23, S(max) = 0.68, H(min) = 0° and H(max) = 50°.

3.3 Morphological Processing

Based on the HSV thresholding, a black and white mask is obtained with all the faces in addition to some artifacts (body parts, background). This mask is then refined through binary morphological operations to reduce the background contribution and remove holes within faces. The image is then eroded to eliminate the small undesired knobs of the region which may affect the template matching (applied later) and cause it to give a false alarm. The eroded image is then dilated to fix some of the undesired effects of the erosion like enlarging of the holes occupying more than 1% of the total size of the image and, refilling the holes which have a size equivalent to or less than the size of the kernel. The last preparation done before the template matching is applying the median filter on the dilated image for removal of the small areas which appear as noise in the skin color image processing output, the source of which comes from the false consideration of some areas from the background as a face region due to the similarity of its color to the skin color space.

3.4 Face Pattern Information

The oval shape of a face can be approximated by an ellipse. Therefore, face detection can be performed by detecting objects with elliptical shape. This can be done based on edges or regions. As a first step, the connected components are located and extracted, and then each connected component is checked whether its shape can be approximated by an ellipse or not.



(a) (b)

Fig. 1: (a) morphological processing, (b) the first connected component.

The connected component is identified by applying the region-growing algorithm described in [23, 24] which depends on finding the 8-connected neighborhoods to a starting pixel which is considered as the first white pixel in the image. This process is iterated until there is no pixel 8-connected to any pixel in the resulting region (see fig. 1).

A hole-filling algorithm is used to refill the gaps inside the region which is not filled by dilation. The connected component C is defined by its center ($\bar{x}$, $\bar{y}$), its orientation θ, and the lengths a and b of its minor and major axis, respectively. The center is computed as the center of mass (centroid) of the

ISSN: 1790-5117 ISBN: 978-960-474-054-3

region. The orientation θ can be computed by elongating the connected component and then finding the orientation of the axis of elongation. This axis corresponds to the least moment of inertia.

$$\theta = \frac{1}{2} arac\ \tan\left(\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}}\right) \quad (1)$$

where

$\mu_{1,1} = \sum x'y' * b(x,y)$, $\mu_{2,0} = \sum (x')^2 * b(x,y)$

$\mu_{0,2} = \sum (y')^2 * b(x,y)$

$x' = x - \bar{x}$, $y' = y - \bar{y}$,

and b(x,y) = 1 if (x,y)∈C, or zero otherwise. For further consideration of C, the value of the orientation, in degrees, with respect to the vertical axis is checked to be in the interval [-45, +45]. Then, the major and minor axes of the connected component are computed by evaluating the minimum and maximum moments of inertia of the connected component with orientation θ:

$$I_{\min} = \sum_{(x,y)\in c} \left[(x-\bar{x})\cos\theta - (y-\bar{y})\sin\theta\right]^2 \quad (2)$$

$$I_{\max} = \sum_{(x,y)\in c} \left[(x-\bar{x})\sin\theta - (y-\bar{y})\cos\theta\right]^2 \quad (3)$$

The lengths of the major and minor axes, a and b respectively, are given by:

$$a = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{\max})^3}{I_{\min}}\right]^{1/8},\ b = \left(\frac{4}{\pi}\right)^{1/4} \left[\frac{(I_{\max})^3}{I_{\min}}\right]^{1/8} \quad (4)$$

Based on the ratio between the major and minor axes, a decision is taken whether the connected component is a candidate face. In this respect, the ratio of the major to minor axes is taken as satisfying the golden ratio:

$$\frac{height}{width} = \frac{1+\sqrt{5}}{2}$$

with an error range from (-0.5, +0.5).

In effect, the distance between the connected component and the best-fit ellipse is determined by counting the holes inside of the ellipse and the points of the connected component that are outside the ellipse. The ratio of the number of false points to the number of points of the interior of the ellipse is calculated and used to determine the ellipses that are good approximation of connected components.

4 Validation

The proposed approach was validated for different configurations via testing images with various sizes, formats (JPEG, GIF, Bitmap), and with different number of people in the images; also, the background varied from simple white background to complex background with different colors. The system recoded an average detection time of 1.7 seconds on a 2.0 GHz processor.

Samples of the experimental results are shown in figures 2 to 6. Figures 2, 3, 4, and 5 display successful counting cases for various image format and sizes. On the contrary, figure 6 illustrates an incorrect counting case. Analysis of the reasons behind unsuccessful detection of the image displayed on figure 6 suggests that dark (black) faces are not correctly detected by the proposed approach. However, it should be noticed that dark faces are correctly detected if there is enough brightness as is illustrated in figures 3 and 4; in such cases, the lighting effect is such that it renders the face lighter than its original dark color and thus can be detected. Many research works is being carried out in this area to arrive at an optimal solution for all cases. Hence, inevitably the images which are taken for a person who is swimming or a person that has a part of his face shaded would be miscounted.

5 Conclusion

In this paper, face detection was investigated using color hue and saturation segmentation as the first cue, followed by template (ellipse) matching. The proposed approach is validated for images with different formats, sizes, number of people, and complexity of the image background. Numerical results suggest that the proposed approach correctly detects and counts non dark faces with reasonable background complexity.

There are several possible extensions of the current work, this includes: (a) Account for dark (black) faces, (b) Gender detection and counting of females vs. males, (c) Color segmentation using adaptive thresholding techniques that would increase both the robustness of larger variations of illumination and the portability between different camera systems., (d)

Account for separation of overlapping faces, such as creating more accurate eye and nose kernel, can possible help with this.

Fig. 2: 633x640 JPEG (22.6 KB) image,
# of people = 2, computed # of people = 2

Fig. 3: 155x260 JPEG (11.2 KB) image,
# of people = 4, computed # of people = 4

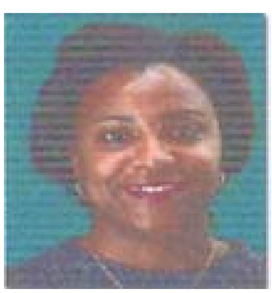


Fig. 4: 200x267 JPEG (6.93 KB) image,
# of people = 1, computed # of people = 1

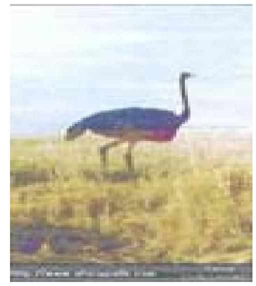


Fig. 5: 270x453 JPEG (34.1 KB) image,
# of people = 0, computed # of people = 0

Fig. 6: 155x260 GIF (11.2 KB) image,
# of people = 5, computed # of people = 4

References:

[1] B. Moghaddam and A. Pentland, “Probabilistic Visual Learning for Object Recognition,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997, pp. 696-710.

[2] H. Verma, P. Sharma, and V.S. Sharma , “A Procedure for Face Detection & Recognition,” Proceedings of the 18th IASTED International Conference: Modelling and Simulation, Montreal, Canada, 2007, pp. 471-476.

[3] M-H Yan, D.J. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey”, IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 24, No. 1, January 2002, pp. 34-58.

[4] V. Vezhnevets, V. Sazonov,and A. Andreeva,” A Survey on Pixel-Based Skin Color Detection Techniques,” Proc. Graphicon, Moscow, Russia, Sep. 2003, pp. 85-92.

[5] G. Burel and D. Caret, “Detection and localization of faces on digital images,” Pattern Recog. Lett. 15, 1994, pp. 963-967.
[6] P.J. L. Van Beek, M.J.T. Reinders, B. Sankur, and J. C. A. Van Der Lubbe, “Semantic segmentation of videophone image sequences,” in Proc. of SPIE Int. Conf. on Visual Communications and Image Processing, 1992, pp. 1182-1193.
[7] R. Brunelli and T. Poggio, “Face recognition: Feature versus templates,” IEEE Trans. Pattern Anal. Mach. Intell. 15, 1993, pp. 1042-1052.
[8] T. Sakai, M. Nagao, and T. Kanade, “Computer analysis and classification of photographs of human faces,” Proc. First USA-Japan Computer Conference, 1972, pp. 2-7.
[9] Craw, H. Ellis, and J.R. Lishman, Automatic extraction of face-feature, Pattern Recog. Lett. Feb. 1987, pp. 183-187.
[10] J. Yang and A. Waibel, “A real-time face tracker,” in IEEE Proc. of the 3rd Workshop on Applications of Computer Vision, Florida, 1996.
[11] Rafael C. Gonzalez Richard E. Woods, Digital Image Processing, 2nd Edition, Prentice Hall, New Jersey, USA, 2002.
[12] H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, “Focus of attention for face and hand gesture recognition using multiple cameras,” in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.
[13] W. Yoo and I. S. Oh, “A fast algorithm for tracking human faces based on chromatic histograms, Pattern Recog. Lett. 20, 1999, pp. 967-978.
[14] K. C. Yow and R. Cipolla, “Feature-based human face detection,” Image Vision Comput. 15(9), 1997.
[15] G. Holst, “Face detection by facets: Combined bottom-up and top-down search using compound templates,” in Proceedings of the 2000 International Conference on Image Processing, 2000, p. TA07.08.
[16] L. Sirovich and M. Kirby, “Low-dimensional procedure for the characterization of human faces,” J. Opt. Soc. Amer. 4, 1987, pp. 519-524.
[17] H. A. Aboalsamh, “Human Face Recognition using Eigen and Fisher Faces,” Egyptian Computer Science Journal, ISSN-1110-2586, vol. 31, no. 1, January 2009.
[18] I.M. Mansour and S.R. Weshah, “A Robust Algorithm for Face Detection in Color Images,” Proceedings of the 16th Danish Conference on Pattern Recognition, 2005.
[19] V. Kumar and T. Poggio, “Learning-based approach to real time tracking and analysis of faces,” in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.
[20] J. Ng and S. Gong, “Performing multi-view face detection and pose estimation using a composite support vector machine across the view sphere,” in Proc. IEEE International Workshop on Recognition, Analysis and Tracking of Faces and Gesture in Real-Time Systems, 1999.
[21] Rong Xiao , Ming-Jing Li, Hong-Jiang Zhang , “Robust multipose face detection in images,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, issue 1, Jan. 2004, pp. 31–41.
[22] N. Farajzadeh and K. Faez, “Hybrid face detection system with robust face and non-face discriminability,” 23rd International Conference on Image and Vision Computing New Zealand, 2008, pp.1-6.
[23] A. S. Al-Humaimeedy, “Automatic Face Detection for Count Estimates,” Master Project in Computer Sciences, College of Computer and Information Sciences, King Saud University, 2003.
[24] Sobottka, K. and Pittas, I., “A novel method for automatic face segmentation, facial feature extraction and tracking,” Signal Processing: Image Communication 12, 1998, pp.263-281.