Accurate extraction of human faces and their components from color digital images based on a hierarchical model

Xi-wen Zhang
Dept. of Digital Media, College of Information Sciences
Beijing Language and Culture University
Beijing 100083, China

Michael R. Lyu
Dept. of Computer Science & Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong SAR, China

***Abstract*-Many human face-based applications depend on accurate extraction of human faces and their components from complex color digital images. Most existing methods are only based on human skin regions and their adjacency relations, but do not utilize their sub-division and grouping at multiple levels. This paper presents a hierarchical model to address this. The model contains pixels, runs, regions, and their respective relations for a digital image. Human skin regions are identified by combining multiple color spaces. Adjacent human skin regions are grouped as a patch. Human face candidates are extracted from each refined human skin patch by un-linking adjacent regions of some regions based on a human facial shape model. Non-human skin patches in a human face candidate are identified as human facial internal component candidates according to the attributes and configurations of human facial components. Each human face candidate is further classified as a human face or not based on its component candidates. An extracted human face provides not only an accurate human facial region but also its components. Finally, this paper demonstrates experimental results of extracting various human faces and their components from three databases, showing that the proposed approach is more accurate and robust than other approaches.**

Keywords-color digital image; human face detection; hierarchy modle; mean shift; region adjacency graph

I. INTRODUCTION

Multiple human faces can appear in a gray or color digital image. Many human face-based applications, such as: human facial expression analysis, human face recognition, human face tracking, and content-based retrieval for digital images and digital videos, require the extraction of human faces and their components. All human faces in a digital image should be accurately and reliably detected, located, extracted and segmented, regardless of their sizes, positions, expressions, orientations, and light illumination conditions. Many methods have already been applied to this problem in recent decades and can deliver excellent results in some cases; however, the problem is still not solved satisfactorily [1, 2].

More and more color digital images with a higher quality are being captured nowadays due to the continual improvement of digital imaging technology. Most approaches to extracting human faces from color digital images are based on a human skin color and identification of a number of human skin regions [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. Most of them cannot accurately extract human faces and their components from color digital images with complex foregrounds and backgrounds because they pay minimal attention to the regions' subdivision and grouping at multiple levels. In practice, accurate extraction of human faces and their components can provide more information, such as human face styles and human facial components' states for analyses and animation. Thus, we propose a novel approach based on a hierarchical model working from pixels to runs to regions and patches and defining their respective relations.

The rest of the paper is organized as follows. Section 2 reviews and analyzes previous work related to human face detection and extraction. Section 3 describes the construction of a hierarchical model for a color digital image. Section 4 presents our technique for extraction of human face candidates. Section 5 presents extraction of human facial component candidates and classification of human face candidates. Section 6 reports experimental results and performance analysis, and Section 7 draws some conclusions.

II. ANALYSIS OF THE RELATED WORK

There have been many investigations into human face detection and extraction from gray digital images [1, 3, 4, 5, 6, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. However, human skin colors have proven to be a more useful and robust cue for human face detection, localization and tracking in color digital images and digital videos [33, 34]. Typically, the process of detecting and extracting human faces from color digital images based on human skin regions is performed in three stages: 1) Identify human skin regions in a color digital image; 2) Find human face candidates from a human skin region or form a human face candidate by combining multiple human skin regions; 3) Decide whether each human face candidate is a real human face.

A. Approaches for identifying a human skin color

Human skin pixels and regions are identified based on the human skin color. The human skin color is calculated from many human skin digital images. The human skin color cluster is fitted to certain geometric shapes. In general, color digital images are stored in a RGB format. It is difficult to identify human skin pixels (regions) in RGB color space because RGB components are influenced by lighting conditions. To circumvent this, many other color spaces can be used instead. Chiang et al. [2] adopt a polynomial model in normalized RG. Hsieh et al. [7] use three rectangles in HS and a threshold of I.

978-1-4244-9306-7/11/$26.00 ©2011 IEEE

Hsu et al. [8] fit an ellipse in C_bC_r. Wang et al. [11] define a rectangle in the normalized RG and a cube in HSV. Garcia et al. [17] directly estimate in YC_bC_r and HSV. Wu et al. [16] construct a fuzzy model based on the perceptually uniform color system of Farnsworth.

B. Approaches to identifying human skin regions

Human skin regions can be derived from pixels or regions. In pixel-based approaches, human skin pixels are first identified according to a color constraint in selected color spaces, and then connected human skin pixels are grouped as a region [7]. Garcia et al. [17] first quantize a color digital image into 16 types of color, and then human skin pixels are identified from the quantized image. Wei et al. [14, 15] process digital images with opening and closing morphological operations to eliminate small and narrow human skin regions, detach weakly linked human skin regions from each other, and smooth human skin regions. Wang et al. [11] use evolutionary agents uniformly distributed in a digital image to detect human skin pixels. Multiple agents form a family corresponding to a human skin region.

In the other approaches, a digital image is first segmented into homogeneous regions, and then human skin regions are identified from them according to a color constraint. Albiot et al. [13] acquire homogeneous regions based on its 2D histogram in C_bC_r. A watershed algorithm is applied to the histogram. Fan et al. [12] use a seeded region growing based on its color edge regions. They take regions having average color similarities above a given threshold as human skin regions.

C. Approaches to identifying human face candidates

Human face candidates result from human skin regions by splitting or merging regions.

In splitting-based approaches, Hsieh et al. [7] separate human facial regions using a clustering-based splitting algorithm. Human face candidates are cropped from these human facial regions based on an elliptical human face model. Wei et al. [14, 15] present an iterative region partitioning procedure for frontal human face candidates.

In the other approaches, Wang et al. [11] first uniformly distribute a number of color-sensitive agents in a digital image to cluster the human skin-like color pixels, and then segment each human face candidate by activating these agents' evolutionary behaviors. Wu et al. [16] compare human skin color regions, and the human hair color regions, with pre-built human head-shape models, using a fuzzy-theory based pattern-matching method. Garcia et al. [17] store all human skin regions in a region adjacency graph. Albiol et al. [13] search each pair-wise merging of regions to find the merged region which best fits a human face model.

The above approaches can provide human face candidates with a coarse scale of polygons. Because of this, the extracted human faces are often not accurate; for example, human faces are often linked with necks. Also, approaches based on searching sub-images can only provide human face candidates of rectangular shape [20, 21, 27, 29, 30, 32].

D. Features to classifying human face candidates

Before a human face candidate is further classified as a human face or not, its features must be extracted.

(1) Geometric features

a) Density, $Den = N_P / A_B$, where N_P is the number of human skin pixels it contains and A_B can be the area of its bounding rectangle (for a frontal face) [12] or fitted ellipse [14, 15].

b) Aspect, $Asp = L_S / L_L$, where L_S and L_L can be the minor and major axis lengths of its fitted ellipse [14, 15], or the width and height of its bounding rectangle (for a frontal face) [12].

c) Compactness [10], $Com = P^2 / 4\pi N_P$, where P is its perimeter.

d) Ellipse fitness error (the Hausdorff distance) [12], $D_{CE} = \max(D(C,E), D(E,C))$, where C represents its boundary points, E represents points on its fitted ellipse, and $D()$ is the maximum of minimum distances between C and E.

(2) Human facial components

Hsieh et al. [7] first extract two eyes and one mouth from an elliptical human face candidate using a local threshold. Wang et al. [11] feed regions to a Back-Propagation neural network to identify eyes. Wei et al. [14] extract facial components with a histogram-based threshold on the component of a human face candidate. Darker regions are identified as one eye (an eye-pair) and a mouth based on their spatial information.

(3) Filtered features

Waring et al. [20] exploit histograms of filtered data of a digital image, where the Gradient, the Laplacian of a Gaussian, and the Gabor filters are used. Huang et al. [19] extract features based on Gabor filters due to their desirable characteristics of spatial localization and orientation selectivity. They design four Gabor filters corresponding to four orientations for extracting human facial features from the local digital image in a sliding window. Wang et al. [11] apply a wavelet-decomposition to a human face candidate to detect its possible human facial features. Viola et al. [28] apply AdaBoost to select features from a set of over-complete Harr wavelet features. Hotta [23] detects view-invariant human faces based on local Principal Component Analysis (PCA) cells.

E. Approaches to classifying human face candidates

Many pattern classification approaches have been applied to the task.

(1) Rules-based approaches

Human face candidates can be classified using rules based on their geometric features and facial components. Chiang et al. [2] detect arbitrarily human faces in color digital images by identifying mouth corners and eyes. Fan et al. [12] identify frontal human faces based on density [0.65, 0.8], aspect [0.4,

0.85], and ellipse fitness error. Sandeep et al. [9] identify a human face according to its density larger than 0.55 and aspect $((1+\sqrt{5})/2)\pm 0.65$. Kuchi et al. [10] use rules with respect to density larger than 0.528, aspect [0.9, 2.10], compactness larger than 0.025, and normalized area larger than 0.35. Wang et al. [11] identify a human face if it contains an eye. Wei et al. [14] identify a human face if it contains an eye, an eye-pair, or a mouth. Park et al. [25] identify human skin and human hair regions based on color, and find human faces from the intersection relationship of their convex-hulls.

Current approaches with geometric features can fail if human faces overlap with glasses, hats, and hair in special styles. Approaches based on components may also fail if the eyes are closed, the mouth is wide open, or if the human face is small.

(2) Approaches based on statistical classification

Rowley et al. [32] detect upright frontal human faces based on a neural network. Feraud et al. [27] detect side view human faces and front view human faces from gray digital images using the Constrained Generative Model. Waring et al. [20] detect human faces using spectral histograms and Support vector machines (SVMs). Heisele et al. [21] detect human faces based on feature reduction and hierarchical classification with SVMs. Schneiderman et al. [30] detect human faces in different poses (frontal, left profile and right profile) with multiple Bayes classifiers. Viola et al. [29] developed a fast frontal human face detection system. A cascade of boosting classifiers is built on a set of over-complete Harr-like features that integrates the feature selection and classifier.

These approaches are fast, fairly robust, and advantageous to find small human faces or human faces in poor-quality digital images over the rule-based approaches. But they usually require many positive and negative examples, are computationally intensive, and cannot handle large variations in human face digital images [19, 20]. Moreover, they cannot provide accurate regions of human faces or their components and are able to present rectangles or ellipses of human faces.

In practice, a color digital image can contain multiple human faces with complex foregrounds and backgrounds.

1) Human faces vary substantially in appearance (orientation, size, and brightness), even within a single image.

2) Human facial components have various states; for example, mouths and eyes are open or closed.

3) Human faces are not complete, usually obscured by non-skin objects, e.g., hair, beards, mustaches, glasses and hats.

4) A human face touches or overlaps other human skin regions, such as ears, neck, and other body parts, or another person's face, hands, arms, etc.

To accurately extract human faces and their components from complex color digital images, we use a hierarchical model containing multiple levels of information.

1) Multiple levels of information, with more contexts, is used to detect and extract human faces and their components.

2) A priori knowledge of human faces is effectively incorporated to improve the reliability and.

3) Each human face and its components are extracted.

III. CONSTRUCTING A HIERARCHICAL MODEL

A color digital image is first segmented using the mean shift algorithm. Horizontally adjacent pixels with the same color form a run. Vertically adjacent runs with the same color form a region. Thus, a hierarchical model contains three levels of information from regions to horizontal runs to pixels, and their respective relations. The model is constructed as a hierarchical structure in which each element consists of its sub-elements, making it convenient to extract information (features) at any level, downward to pixels and upward to regions.

A. Segmenting digital images using the mean shift algorithm

A color digital image can be segmented using many approaches [35], e.g., the region growing approach [36], the watershed algorithm [37], the level set approach [38], and the mean shift algorithm [39]. The mean shift algorithm is employed in our work because of its performance and its relative freedom from specifying an expected number of regions. The bandwidths H_s in the color domain and H_r in the range domain are set at 5 and 7, respectively, and the size of the minimum region is set at 49 pixels. These parameters were derived through experiments. A color digital image with multiple human faces is shown in Fig. 1(a), and its segmented version is shown in Fig. 1(b).

(a) An original color digital image.

(b) Its segmented version.

(c) Human skin patches in blue contours and their regions in green.

(d) Refined human skin patches and their regions.

(e) Human face candidates are extracted in blue contours.

(f) Human faces and their components are extracted.

Figure 1. Human faces and their components are extracted.

B. Constructing a run matrix for a segmented digital image

A segmented digital image is first represented as a run matrix. A run is described using its starting and end x coordinates (x_S, x_E), y, as well as its color value (co). The runs are visited in the following sequence: from top to bottom for a digital image, and from left to right along the horizontal direction for each row.

The relations between adjacent runs (R_A and R_B) can be classified into four types according to their relative positions:

1) R_A is the top neighbor of R_B if $y_A - y_B = -1$ and they overlap in horizon.

2) R_B is the bottom neighbor of R_A if R_A is the top one of R_B.

3) R_C is the left neighbor of R_D if $x_{EC} - x_{SD} = -1$ and $y_A = y_B$.

4) R_D is the right neighbor of R_C if R_C is the left neighbor of R_D.

The first two types of relations are for neighboring rows' overlapped runs, and the other relations are for adjacent columns' runs.

Runs form a row of runs if they have the same y. In the same row, a run is the left neighbor of the next run, and the next run is its right neighbor. Runs in a row are top (bottom) neighbors of a run in the next (previous) row if they are overlapped. A run matrix consists of run rows. It is obvious that a matrix of runs represents not only runs themselves but also their adjacent runs in four directions.

C. Constructing a region adjacent graph

A region adjacent graph (RAG) has been used for enhancing various digital image segmentation algorithms [40]. A RAG for a digital image provides not only a spatial view at the region level, but also the higher-order connectivity view. Hence, a RAG represents not only the local region properties but also spatial relationships between neighboring regions. Thus, a RAG is used to represent a digital image at the region level.

Based on the run matrix constructed above, the higher level's representation of a segmented digital image can be obtained by aggregating related runs into regions. Related runs are identified via the matrix of runs, and each group of related runs comprises a region. These related runs must satisfy the following requirements.

1) They have the same color.

2) They are vertically adjacent to each other.

3) Their overlap length is larger than the half of the shorter run's length.

Adjacent regions of a region can be easily determined in terms of the relations between their runs: if a run of one region is adjacent to a run of another region, the two regions are adjacent to each other. Regions and their adjacent regions form a region adjacency graph, in which nodes are regions, and edges are adjacent relations of regions.

A region can be subdivided into multiple sub-regions with more homogenous colors by splitting its runs. If the deviation color value of a pixel in a run is larger than a threshold, then

the run is split at the pixel. The vertically adjacent runs form a sub-region. Thus, the original region changes into more regions. Using this approach, a region containing a neck and a chin of a human face can be split.

Adjacent regions satisfying certain constraints can be grouped as a patch. A patch can be characterized using features mentioned in Section 2.4 and the following features. To accurately compute the attributes of a patch, we use a convex hull [41] as its bounding shape.

1) Smoothness, $Smoo = P_B - P_C$, where P_B, P_C are the perimeter of its boundary polygons and convex hulls, respectively.

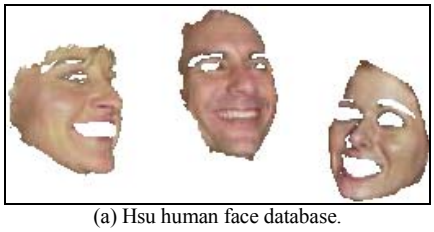
2)Homogeneity, $Homo = \sum_{i=1}^{N}\sum_{j=1}^{N}(P(i,j)/(1+|i-j|))$, where $P(i,j)$ is the color of a pixel.

IV. Identification of Human Face Candidates Based on a Human Facial Geometry Model

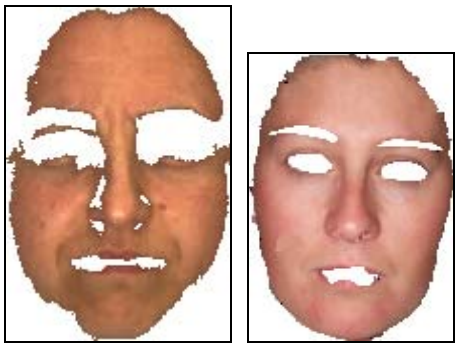
Human skin regions are first identified by combining multiple color spaces, and then adjacent human skin regions are grouped as a patch. Non-human skin pixels of each human skin patch are removed via an adaptive threshold. Human face candidates are then extracted from each human skin patch by un-linking adjacent regions of some regions based on a human facial geometry model.

A. Identifying and grouping of human skin regions

Existing human skin colors are modeled from pixels [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 33]. It can save much time and increase tolerance of noisy pixels if we identify a human skin color based on regions. Human face skin regions are manually cropped from databases [42, 43, 44] containing human faces of different races and ages, under various lighting conditions; some samples are shown in Fig. 2. They are segmented into homogenous regions using the mean shift algorithm. These regions' colors (red points) are modeled as green convex hulls in normalized RG, HS, and C_bC_r, as shown in Fig. 3. V and Y are separated from HS and C_bC_r, respectively. We can see that each human skin color cluster is not a convex hull with a smaller fitting error. Thus we use their interaction to identify human skin regions. This is to extract all human face skin regions as accurately as possible at this initial stage. The regions will be refined in the following stage using more accurate local information. Human face skin regions are bounded by green boundaries shown in Fig. 1(c), and regions with yellow boundaries are non-human skins.

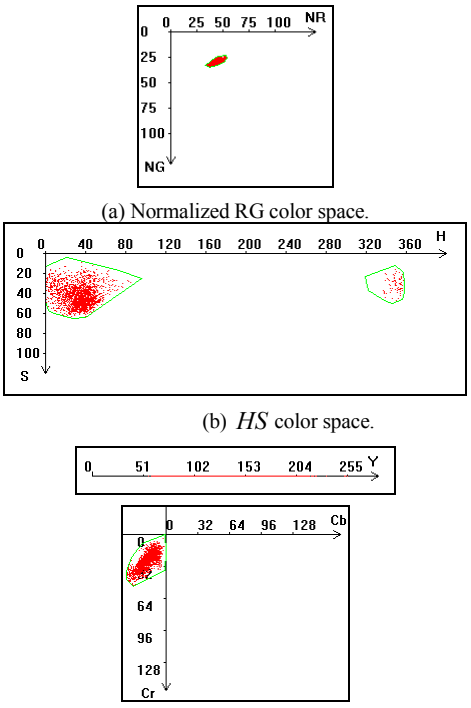


(a) Hsu human face database.



(b) AR face database. (c) Caltech face database.

Figure 2. Human face skins in three databases.



(a) Normalized RG color space.

(b) HS color space.

(c) C_bC_r color space.

Figure 3. Clusters of human face skin regions in color spaces.

Adjacent human skin regions are grouped as a patch. A human skin patch can contain one human face, several human faces, and/or other human skin regions. Fig. 1(c) shows human skin patches with blue boundaries and numbers. Human faces in different digital images are different in color, and so are human faces within a single digital image. However, human skin regions within a single human face have similar colors. Thus, non-human skin pixels of each human skin patch are removed via an adaptive threshold from its original pixels. Normalized red values (NRV) of pixels in a human skin patch can be classified into two types: those of human skins and those of non-human skins. In order to distinguish them, a histogram is first constructed, and an optimal distinction threshold [45] from it can be identified that minimizes the NRV classification error. Some pixels of a human skin patch are removed from their corresponding runs and regions if their NRVs are less than the threshold. The refined human skin patches are shown in Fig. 1(d).

B. Extraction of human facial candidates from each human skin patch

A complete human face consists of a skin and components containing two eyebrows, two eyes (open or closed), two nostrils, a mouth (with teeth and/or tongue for an open mouth), and two ears. Human facial internal components (e.g. eyebrows, eyes, mouth, nostrils) are not human skin regions. Due to various hairstyles and overlapped objects, a human face skin has various shapes. However, the density of a human face (except its ears) is close to 1, using its convex hull. Thus, if a human skin patch satisfies the following conditions then it is a human face candidate: 1) Its smoothness is smaller than 20 pixels; 2) Its aspect ranges from 0.5 to 2.0; 3) The maximum ellipse fitting error of its convex hull is smaller than 7.

The other human skin patches may contain more than one human face and/or other human skin regions. We can see that, from Fig. 1(d), if certain regions are removed from human skin patches (in the cases of patches 0, 1, 3, 5, 6, 7, 8, and 9), complete and accurate human face skins are attained. Human face candidates contained in a human skin patch are sub-patches with a relatively high human facial geometry confidence and satisfying the above conditions. We define the confidence of human facial geometry as $S_{FC} = Smoo + D_{CE}$, where $Smoo$ and D_{CE} are smoothness and ellipse fitting error of a sub-patch, respectively. Regions in a human skin patch are connected in multiple ways, and most of their connections can be removed. We first unlink adjacent regions of some regions if they keep a human skin patch unchanged. This can reduce further combinations of removed regions. Then, human face candidates are extracted from a human skin patch by un-linking adjacent regions of some regions. Many human face candidates cannot be accurately extracted if regions are removed one by one. Therefore, we have to consider removing multiple regions at one time. Fortunately, adjacent regions of a few regions need to be removed in order to attain a refined human skin-patch (human face candidates). Human face candidates are found from each human skin-patch using the following steps:

1) Generate $\sum_{M=0}^{M-1} C_N^M$ strings. Each string has N characters. Each character within the strings is 1 or 0. N is denoted as the number of edges. There are at most M 0's in a string, such that $3 < M > N/5$.

2) If sub-patches corresponding to a string have the maximum sum of S_{FC} and satisfy the conditions in the above paragraph, then they are set as human face candidates.

Fig. 1(e) shows extracted face candidates from the human skin patches shown in Fig. 1(d). Human face candidates are bounded in blue contours and labeled by blue numbers.

V. Classification of Human Face Candidates Based on Human Facial Internal Component Candidates

A human face can be identified according to its components. Thus, we extract human facial component candidates from a human face candidate and identify a real human face using them. Human facial internal components are non-human skin, and are identified from non-human skin patches contained in the convex hull of a human face candidate. Regions of an eye and an eyebrow may connect while they have blur boundary. A pair of nostrils can be a single region, or their regions can be connected. Regions for a mouth can connect a mouth's corner regions. So, each non-human skin patch is split using the approach in Section 4.2 because human facial components are also compact. Rules for identifying a human facial component are as follows: 1) Its density is bigger than 0.9; 2) Its aspect ranges from 0.25 to 4.0; 3) The maximum ellipse fitting error is smaller than 4.

A closed eye is similar to an eyebrow in texture. An open eye is similar to an eyebrow in shape, but their textures are different. A mouth, when closed or slightly open, is similar to an open eye in shape, but their components are different. There are at most two eyebrows, two eyes and two nostrils in a human face candidate. Each human facial component has its own characteristics of shape, texture, and composition. Moreover, components in a human face have many topological relations. Thus, we extract human facial components from a human face candidate based on shapes, textures, compositions, and spatial relations of its non- human skin patches.

A non-human skin patch in a human face candidate can be identified as an eyebrow, an eye, a nostril, a mouth candidate, or a noisy patch. A human face can appear in a frontal view, side view, top-half (a neckerchief covering the mouth, or even the nose), or bottom-half (hair covering the forehead and a pair of glasses separates the human face into two parts). We collect four types of human faces and their variations as samples. The minimum and maximum thresholds in extracting component candidates from each human face candidate are decided through many samples, as shown in Table I. Thresholds are adaptive, varying amongst the human face candidates.

TABLE I. RULES TO IDENTIFY HUMAN FACIAL COMPONENTS

Characteristics / Components	L_{LC}/L_S	L_{SC}/L_L	Den	$Homo$
Eyebrow	>0.14 <0.35	>0.04 <0.12	>0.75	>0.85
Eye	>0.13 <0.24	>0.03 <0.13	>0.85	>0.75
Nostril	>0.04 <0.15	>0.02 <0.10	>0.90	>0.90
Mouth	>0.15 <0.50	>0.05 <0.25	>0.80	>0.75

According to the above thresholds, a single non-human skin patch in a human face candidate can be identified as a candidate for an eyebrow, an eye, and a mouth. Candidates having more than one possible type are re-identified. Topology relations of human facial component candidates in a human face candidate are utilized to identify real human facial components. Spatial relations among components are learned from samples. The following features are extracted, as shown in Table II.

Components of each human face sample are extracted manually. Features of components of each human face are computed and stored as human facial components in the

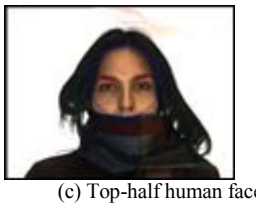
database. Some samples of human facial components are shown in Fig. 4. Extracted features are shown in Table III.

TABLE II. SPATIAL FEATURES OF FACIAL COMPONENTS

Spatial features	Distances	L_S and L_L	Configurations
$R_{Eb-Eb} = D_{Eb-Eb} / L_S$	D_{Eb-Eb} is the distance between two eyebrows' centers.	L_S is the short axis length of a face candidate's convex hull.	A horizontal configuration relates to pairs of eyebrows, eyes, and nostrils.
$R_{E-E} = D_{E-E} / L_S$	D_{E-E} is the distance between two eyes' centers.		
$R_{N-N} = D_{N-N} / L_S$	D_{N-N} is the distance between two nostrils' centers.		
$R_{Eb-E} = D_{Eb-E} / L_L$	D_{Eb-E} is the distance between centers of an eyebrow and an eye and it is smaller than half of the eye's length.	L_L is the long axis length of a face candidate's convex hull.	A vertical configuration represents relations between an eyebrow, an eye, a nostril, and a mouth.
$R_{Eb-N} = D_{Eb-N} / L_L$	D_{Eb-N} is the distance between centers of an eyebrow and a nostril.		
$R_{Eb-Mo} = D_{Eb-Mo} / L_L$	D_{Eb-Mo} is the distance between centers of an eyebrow and a mouth.		
$R_{E-N} = D_{E-N} / L_L$	D_{E-N} is the distance between centers of an eye and a nostril.		A vertical configuration denotes relations between an eye, a nostril, and a mouth.
$R_{E-Mo} = D_{E-Mo} / L_L$	D_{E-Mo} is the distance between centers of an eye and a mouth.		
$R_{N-Mo} = D_{N-Mo} / L_L$	D_{N-Mo} is the distance between centers of a nostril and a mouth.		A vertical configuration is for relations between a nostril, a mustache, and a mouth.


(a) Frontal human faces.


(b) Side human faces.


(c) Top-half human faces.


(d) Bottom human faces.

Figure 4. Human facial components.

TABLE III. SPATIAL FEATURES

Human faces / Characteristics	Frontal faces	Side faces	Top-half faces	Bottom-half faces
R_{Eb-Eb}	>0.45 <0.50	Null	>0.40 <0.50	Null
R_{E-E}	>0.44 <0.49	Null	>0.39 <0.49	Null
R_{N-N}	>0.11 <0.15	Null	>0.11 <0.15	>0.13 <0.15
R_{Eb-E}	>0.07 <0.12	>0.11 <0.13	>0.14 <0.24	Null
R_{Eb-N}	>0.27 <0.39	>0.32 <0.36	>0.47 <0.66	Null
R_{Eb-Mo}	>0.42 <0.56	>0.44 <0.48	Null	Null
R_{E-N}	>0.20 <0.34	>0.26 <0.32	>0.31 <0.49	Null
R_{E-Mo}	>0.34 <0.51	>0.36 <0.43	Null	Null
R_{N-Mo}	>0.14 <0.19	>0.15 <0.18	Null	>0.26 <0.36

A non-human skin patch of each human face candidate may be identified as multiple human facial components using the above rules without considering their spatial relations. Thus, human facial component candidates of a human face candidate have many configurations. The best configuration of component candidates of a human face candidate is identified using a dynamic programming technique based on the database of human facial components' configuration, as shown in Fig. 5. To attain the configuration score of a human face's components, the attributes listed in Table II are first calculated; the score is equal to the absolute value of the difference between the computed attribute values and the corresponding attribute values list in Table III. Referenced values must belong to the same human face model. Human facial component candidates extracted from human face candidates are shown in Fig. 1(e); eyebrow candidates are shown in red contours, eye candidates in brown contours, nostril candidates in black contours, and mouth candidates in pink contours.

1) Set the first human facial component candidate as the previous configuration of human facial components (PC).
2) Identify the best configuration for component candidates of a human face candidate.
For i = 1 to N-1 (N is the number of human facial component candidates in a human face candidate)
{
  Calculate scores for the ith human facial component candidate with the PC, and set the configuration with the maximum Sc as the PC.
}
3) Human facial components are identified according to their best configuration.

Figure 5. Human facial components are identified.

A human face candidate is identified as a human face if it contains: 1) a mouth and a nostril-pair or an eye-pair or a pair

of eyebrows (frontal human faces), 2) a mouth (bottom-half human faces), 3) an eye-pair (top-half human faces), 4) a mouth and an eye (side view human faces). Extracted human faces and their components are shown in Fig. 1(f). The nodes in the second top level are a human face skin, and human facial components. The human facial components contain eyebrows, eyes, nostrils, and a mouth.

VI. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSES

Based on the proposed approach, we have developed a software prototype in Visual C++ R7.0. This section presents performance evaluation and comparison based on experimental results and ground-truth data.

A. Experimental results

To evaluate the performance of the proposed method, we have conducted experiments on three human face databases [42, 43, 44], whose information is shown in Table IV. The human faces in them appear in simple background [42], natural background [43], or complex foreground and background [44]. More experimental results are shown in Fig. 6 to illustrate the effectiveness of our approach. We can see that many kinds of human faces are extracted correctly and accurately.

TABLE IV. AVERAGE PERFORMANCE IN CORRECTNESS

Face database		Detected object	Number	Precision (%)	Recall (%)
AR CD-8		Face	273	88	87
		Eyebrow	420	85	86
		Eye	420	86	87
		Mouth	210	85	86
Caltech		Face	447	91	92
		Eyebrow	893	90	89
		Eye	893	89	88
		Mouth	447	88	89
Hsu	012	Face	53	89	89
		Eyebrow	104	86	85
		Eye	104	85	86
		Mouth	53	87	86
	043	Face	123	88	88
		Eyebrow	230	89	89
		Eye	230	86	86
		Mouth	115	87	88
	327	Face	621	88	87
		Eyebrow	1125	84	83
		Eye	1145	82	81
		Mouth	587	85	85

(a) AR. (b) Caltech. (c) Hsu.

Figure 6. Some human faces from three databases are extracted and segmented.

The performance of our approach is evaluated for extracted human faces and their components in two aspects: their correctness and region accuracy. To make a quantitative evaluation on the extracted results, the ground-truth data are used as a reference. As the ground-truth data cannot be obtained by any automatic processing, professional engineers manually generated them. The precision rate and the recall rate [2, 18] are used to measure the average performance of detected objects in terms of correctness. The average correctness performance of faces and their components is given in Table IV.

We further evaluate the region accuracy of extracted objects by employing two well-known digital image segmentation evaluation protocols that ranked high in Zhang's survey [45]: Relative Ultimate Measurement Accuracy for Area (AA) [46] and Probability of Error (PE) [47]. AA is defined as $AA = |R_A - S_A| / R_A \times 100\%$, where R_A is the area of the ground-truth object and S_A is the area of the extracted object. Thus, $RUMA_A$ indicates the relative percentage of the area discrepancy. The smaller the AA is, the higher the extraction accuracy is. PE is defined as $PE = P(O) \times P(B|O) + P(B) \times P(O|B)$, where $P(B|O)$ is the probability of erroneously classifying objects as background, $P(O|B)$ is the probability of erroneously classifying background as objects, and $P(O)$ and $P(B)$ are a priori probabilities of objects and background existing in digital images. Also, the smaller the PE is, the higher the extraction accuracy is.

The evaluation results of the pixel discrepancy of extracted human faces and their components compared with the ground-truth data are given in Table V. Table V demonstrates that the extraction discrepancy of the proposed approach is high. Actually, the average extraction discrepancies in AA and in PE are below 13% and below 0.15 respectively.

TABLE V. EVALUATION OF EXTRACTION DISCREPANCY

Face database	Face		Eyebrow		Eye		Mouth	
	AA	PE	AA	PE	AA	PE	AA	PE
AR	7%	0.10	10%	0.13	9%	0.80	8%	0.11
Caltech	7%	0.09	11%	0.12	9%	0.11	8%	0.11
Hsu	8%	0.11	13%	0.15	11%	0.12	9%	0.13

Processing times of human facial candidates and faces are tested on a PC with PIII 1.6 GHz CPU and 512M RAM, giving a maximum time of 1.87 seconds, and an average time of 1.45 seconds.

B. Comparison with related work

There are many approaches to extracting human faces and their components from color digital images. In this section, we compare them with respect to four characteristics: computational cost, human facial candidates, human face, and human facial component, as listed in Table VI.

Many human face databases have been used for quantitative testing of the various approaches listed above [2, 7, 8, 11, 16, 17, 18, 19]. We were able to obtain some of them

[2, 8] and other human face databases [43, 44]. Others' performance evaluations in terms of correctness on these two human face databases [2, 8] are listed in Table VII. From Table IV and VII, we can see that our approach is close to or better than others in terms of precision and recall rates. However, others' work does not evaluate their results' accuracy. Our investigations show that they provide lower accuracy; this is because they extract human face regions in polygons with coarse scales, yielding an average AA greater than 17% and PE greater than 0.16. Our approach can provide accurate human faces and their components because we consider the splitting and merging of regions at multiple levels.

TABLE VI. COMPARISON BETWEEN TWO APPROACHES

Characteristic / Approach	Human facial candidate	Human Face	Facial component	Computational cost
Learning-based	A sub-image	Circle Ellipse Rectangle	No	High
Splitting of a skin region	A skin region	Polygon with lower scales	No	Medium
Merging of skin regions	Skin regions and their adjacency relations	Polygon with a higher scale	No	Medium
Our approach	pixels, horizontal runs, and regions	Polygon with a higher scale	Yes	Low

TABLE VII. PERFORMANCE OF OTHERS' WORK

Tester	Face database		Precision (%)	Recall (%)	AA	PE
	Name	Total faces				
Chiang et al. [2]	**AR**	**945**	92.98	92.48	16%	0.15
Hsu et al. [8]	**Hsu**	**684**	80.35	89.59	17%	0.16

C. Error analyses

There are many errors during extracting and segmenting human faces. They may occur while identifying human skin regions, human face candidates, human facial components, and human faces. Courses are analyzed as follows.

1) Identification of human facial candidates

Some human facial candidates are missing because their human skin regions have special colors. The model of human skin colors should be furthered. A human skin patch is wrongly identified as a human facial candidate when its shape is similar to that of a face. But we can identify wrong human facial candidates according to their non-human skin patches.

2) Extraction of human faces and their components

Missing of human facial candidates must result in the failure to extract human faces and their components. Some missing human faces are resulted from the wrong identification of human facial component candidates.

D. Discussion

From the above experimental results and our performance analyses and comparisons, it can be concluded that the proposed approach has three major advantages:

1) The hierarchical model can provide more context and evidence for identifying human facial candidates and human faces.

2) More sources of information are exploited to extract human faces, including color, shape, as well as components and their topological information.

3) An extracted human face is represented as a graph, which stores all information extracted and segmented.

Our results show that the proposed approach can achieve satisfactory results.

VII. CONCLUSIONS

This paper proposes the extraction of human faces and their components from complex color digital images based on a hierarchical model. Human faces with complex foregrounds and backgrounds can be processed, thanks to the exploitation of multiple levels of information. Extracted and segmented human faces are represented as a face hierarchy.

Our software implementation of the proposed approach has been tested using a large number of digital images containing human faces. Their performance analyses are reported, including the test results and the performance evaluation. The results confirm that the proposed approach is more effective and robust than other approaches.

ACKNOWLEDGMENT

The work in this paper was substantially supported by grants from the National Natural Science Foundation of P.R. China and the Microsoft Asia Research (Grant No. 60970158), and the Shun Hing Institute of Advanced Engineering (SHIAE) of The Chinese University of Hong Kong.

REFERENCES

[1] Peichung Shih, Chengjun Liu, "Face detection using discriminating feature analysis and Support Vector Machine", Pattern Recognition, 2006, 39(2): 260-276.

[2] Cheng-Chin Chiang, Chi-Jang Huang, "A robust method for detecting arbitrarily tilted human faces in color images", Pattern Recognition Letters, 2005, 26(16): 2518-2536.

[3] Seong G. Kong, Jingu Heo, Besma R. Abidi, Joonki Paik, Mongi A. Abidi, "Recent advances in visual and infrared face recognition-a review", Computer Vision and Image Understanding, 2005, 97(1): 103-135.

[4] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, "Face recognition: A literature survey", ACM Computing Surveys, 2003, 35(4): 399-458.

[5] Ming-Hsuan Yang, David Kriegman, Narendra Ahuja, "Detecting faces in images: A survey", IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, 24(1): 34-58.

[6] E. Hjelmas, B.K. Low, "Face detection: A survey", Computer Vision and Image Understanding 2001, 83(3): 236-274.

[7] Ing-Sheen Hsieh, Kuo-Chin Fan, Chiunhsiun Lin, "A statistic approach to the detection of human faces in color nature scene", Pattern Recognition, 2002, 35(7): 1583-1596.

[8] Rein-Lien Hsu, Mohamed Abdel-Mottaleb, Anil K. Jain, "Face detection in color images", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 696-706.

[9] K. Sandeep, A.N. Rajagopalan, "Human face detection in cluttered color images using skin color and edge information", The Third Indian Conference on Computer Vision, Graphics and Image Processing, 2002.

[10] Prem Kuchi, Prasad Gabbur, P. Subbanna Bhat, Sumam David S., "Human face detection and tracking using skin color modeling and connected component operators", IETE Journal of Research, 2002, 38(3&4): 289-293.

[11] Yanjiang Wang, Baozong Yuan, "A novel approach for human face detection from color images under complex background", Pattern Recognition, 2001, 34(10): 1983-1992.

[12] Jianping Fan, David K. Y. Yau, Ahmed. K. Elmagarmid, Walid G. Aref, "Automatic image segmentation by integrating color-edge extraction and seeded region growing", IEEE Transactions on Image Processing, 2001, 10(10): 1454-1466.

[13] Alberto Albiol, Luis Torres, Charles A. Bouman, Edwar J. Delp, "A simple and efficient face detection algorithm for video database applications", IEEE International Conference on Image Processing, Vancouver, Canada, September 10-13, 2000.

[14] Gang Wei, Ishwar K. Sethi, "Omni-face detection for video/image content description", Proceedings of the 2000 ACM workshops on Multimedia, Los Angeles, California, United States, 2000: 185-189.

[15] Gang Wei, Ishwar K. Sethi, "Face detection for image annotation", Pattern Recognition Letters, 1999, 20(9): 1313-1321.

[16] Haiyuan Wu, Qian Chen, Masahiko Yachida, "Face detection from color images using a fuzzy pattern matching method", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(6): 557-563.

[17] Christophe Garcia, Georgios Tzirtias, "Face detection using quantized skin color regions merging and wavelet packet analysis", IEEE Transactions on Multimedia, 1999, 1(3): 264-277.

[18] J. Cai and A. Goshtasby, "Detecting human faces in color images", Image and Vision Computing, 1999, 18(1): 63-75.

[19] Lin-Lin Huang, Akinobu Shimizu, "Robust face detection using Gabor filter features", Pattern Recognition Letters, 2005, 26(11): 1641-1649.

[20] Christopher A. Waring, Xiuwen Liu, "Face detection using spectral histograms and SVMs", IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2005, 35(3): 467-476.

[21] Bernd Heisele, Thomas Serre, Sam Prentice, "Hierarchical classification and feature reduction for fast face detection with support vector machines", Pattern Recognition, 2003, 36(9): 2007-2017.

[22] Jianxin Wu, Zhi-Hua Zhou, "Efficient face candidate selector for face detection", Pattern Recognition, 2003, 36(5): 1175-1186.

[23] Kazuhiro Hotta, "View-Invariant Face detection method based on local PCA cells", Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP'03).

[24] Olugbenga Ayinde, Yee-Hong Yang, "Region-based face detection", Pattern Recognition, 2002, 35(10): 2095-2107.

[25] Minsick Park, Chang-Woo Park, "Algorithm for detecting human faces based on convex-hull", Optics Express, 2002, 10(6): 274-279.

[26] Chiunhsiun Lin, Kuo-Chin Fan, "Triangle-based approach to the detection of human face", Pattern Recognition, 2001, 34(6): 1271-1284.

[27] R. Fe|raud, O. Bernier, J.-E. Viallet, M. Collobert, "A fast and accurate face detector based on neural networks", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(1): 42-53.

[28] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[29] P. Viola, M. Jones, "Robust real time object detection", IEEE ICCV Workshop on Statistical and Computational Theories of Vision, Vancouver, Canada, July 13, 2001.

[30] H. Schneiderman, T. Kanade, "A statistical method for 3D object detection applied to faces and cars", International Conference on Computer Vision, 2000.

[31] Jianming Hu, Hong Yan, "Locating head and face boundaries for head-shoulder images", Pattern Recognition, 1999, 32(8): 1317-1333.

[32] Henry A. Rowley, Shumeet Baluja, Takeo Kanade, "Neural Network-based face detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 23-38.

[33] Vezhnevets V., Sazonov V., Andreeva A., "A survey on pixel-based skin color detection techniques", Proceeding of Graphicon-2003, Moscow, Russia, September 2003: 85-92.

[34] Son Lam Phung, Abdesselam Bouzerdoum, "Skin segmentation using color pixel classification: analysis and comparison", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(1): 148-154.

[35] H. D. Cheng, X. H. Jiang, Y. Sun, J. Wang, "Color image segmentation: advances and prospects", Pattern Recognition, 2001, 34(12): 2259-2281.

[36] S. A. Hijjatoleslami, J.Kittler, "Region growing: A new approach", IEEE Transaction on Image Processing, 1998, 7: 1079-1084.

[37] L. Vincent, P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations", IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(6): 583-598.

[38] L. Vese, T. Chan, "A multiphase level set framework for image segmentation using the Mumford and Shah model", International Journal of Computer Vision, 2002, 50(3): 271-293.

[39] Dorin Comaniciu, Peter Meer, "Mean shift: A robust approach toward feature space analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(5): 603-619.

[40] Alain Tremeau, Philippe Colantoni, "Regions adjacency graph applied to color image segmentation", IEEE Transactions on Image Progressing, 2000, 9(4): 735-744.

[41] Joseph O'Rourke, "Chapter. 3 Convex hulls in 2D", Computational Geometry in C (2nd Edition), 1998.

[42] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report #24, June 1998, http://rvl1.ecn.purdue.edu/~aleix/ar.html.

[43] Caltech face database, http://www.vision.caltech.edu/Image_Datasets/faces/faces.tar.

[44] Hsu et al. face database, http://www.cse.msu.edu/~hsureinl/facloc/index_facloc.db.html.

[45] Milan Sonka, Vaclav Hlavac, and Roger Boyle, Image Processing, Analysis, and Machine Vision, Second Edition, Brooks/Cole, a division of Thomson Aisa Pte Led, United States America, 1998: 128-130.

[46] Y.J. Zhang, "A survey on evaluation methods for image segmentation", Pattern Recognition, 1996, 29(8): 1335-1346.

[47] Y.J. Zhang, "Evaluation and comparison of different segmentation algorithms", Pattern Recognition Letter, 1997, 18(10): 963-974.