



## DEWA: A Multiaspect Approach for Multiple Face Detection in Complex Scene Digital Image

Setiawan Hadi<sup>1</sup>, Adang S. Ahmad<sup>2</sup>, Iping Supriana Suwardi<sup>2</sup> & Farid Wazdi<sup>2</sup>

<sup>1</sup> Mathematics Department, Padjadjaran University, Indonesia

<sup>2</sup> STEI, Institute of Technology Bandung, Indonesia

**Abstract.** A new approach for detecting faces in a digital image with unconstrained background has been developed. The approach is composed of three phases: segmentation phase, filtering phase and localization phase. In the segmentation phase, we utilized both training and non-training methods, which are implemented in user selectable color space. In the filtering phase, Minkowski addition-based objects removal has been used for image cleaning. In the last phase, an image processing method and a data mining method are employed for grouping and localizing objects, combined with geometric-based image analysis. Several experiments have been conducted using our special face database that consists of simple objects and complex objects. The experiment results demonstrated that the detection accuracy is around 90% and the detection speed is less than 1 second in average.

**Keywords:** *complex background; digital image; face detection; facial image; multiaspect approach; multiple face.*

### 1 Introduction

Information technology development and progress have attracted researchers to bring conceptual ideas of image processing and computer vision into reality. This condition also becomes a driving-force factor for face detection research. In this research, numerous methods have been proposed from various points of view [1]. However, most of them were evaluated using simple uniform scene particularly for detecting single face in image. Only a handful of research, if any, addresses the problem of detecting multiple faces in complex image with unconstrained and various backgrounds. This motivates our research.

### 2 Problem Description of Detecting Faces in Digital Image

A face detection system has to be able to extract the information contained in an image and store it in the form of graphical and descriptive data. Graphical data corresponds to face objects. Descriptive data describes the size, location and orientation of each object (i.e., face) in the image. Conceptually, face detection is a two-class pattern classification problem whose output is the information about the existence of faces in the image.

Face detection is an important problem with respect to the more general context in face recognition. It can be considered as the pre-processing step in a problem that utilizes faces as the objects of interest. The main challenge that has motivated our research is in the face variability (extra-personal as well as intra-personal variability). Extra-personal variability appears on the face objects with different races, gender and other genetics variations. Intra-personal variability related to face characteristics in image such as pose, structural component, facial expressions, object occlusion, image inline orientation and acquisition condition [1], [2].

A face detection system consists of methods for identifying and determining the position of human faces in a digital image, which has to be independent from position, size, orientation, image illumination and other constraints. This condition is called invariant factor of face. This mechanism is relatively easy to perform by God's grace of human biological vision. As an illustration, Figure 1 shows an image with many face objects and various sizes. Face objects can be easily identified by human, but difficult to be identified by a machine.



**Figure 1** Easy or difficult in detecting faces? (profs and students of faculty of Computer Science Univ. of New Brunswick, class'94).

Numerous approaches have been proposed for face detection [1], [3], [4], [5]. However, most of them are focused on a single face detection with uniform and

simple background. Yang, Kriegman dan Ahuja [1] have classified the face detection methods based on four categories: (1) Knowledge-based methods (KBM), (2) Feature-based methods (FBM), (3) Template-based methods (TBM) and (4) Appearance-based methods (APM). KBM methods encode human knowledge of what constitutes a typical face, which is the relationship between facial features. FBM methods aim to find structural features of a face that exist even when the pose, viewpoint, or lighting conditions vary. TBM methods utilize several standard patterns stored to describe the face as a whole or the facial features separately. APM methods focus on the models (or templates) learned from a set of training. These models capture the representative variability of facial appearance.

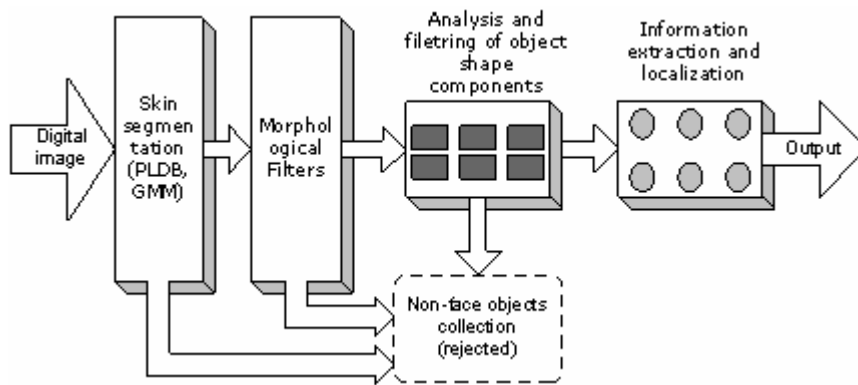
Two knowledge-based methods have been proposed. Yang [2] utilized multi-resolution focus-of-attention approach. It consists of three levels. Level 1 (lowest resolution) applies the rule “the center part of the face has 4 cells with a basically uniform intensity” to search for candidates. Level 2 uses local histogram equalization followed by edge detection. Level 3 searches for eye and mouth features for validation. Kotropoulos and Pitas [6] used horizontal/vertical projection to search for candidates and using eyebrow/eyes, nostrils/nose for validation. Their methods have a drawback in that it is difficult to detect multiple people in complex background. KBM is easy to come up with simple rules to describe the features of a face and their relationships. It is based on the coded rules; facial features in an input image are extracted first, and face candidates are identified. KBM works well for face localization in uncluttered background. However, translating human knowledge into rules precisely is not an easy task; detailed rules fail to detect faces and general rules may find many false positives. In addition, it is difficult to extend this approach to detect faces in different poses and implausible to enumerate all the possible cases.

Feature-based methods are considered as a bottom-up approach. They detect facial features (eyes, nose, mouth, etc) first. Facial features such as edge, intensity, shape, texture, color are exposed. FBM aims to detect invariant features. The methods group features into candidates and verify them. The drawback of FBM is that it is difficult to locate facial features due to several corruptions (illumination, noise and occlusion). Features in complex background are also not easy to detect.

Face detection has been developed for many years but several problems are still unresolved, especially in the context of multiple face detection for face recognition [7]. It inspired many researchers to explore and to find the solution for the problem.

### 3 Proposed Approach for Detecting Face in Digital Image

Our proposed approach for detecting faces in image consists of three phases, which is a combination of KBM and FBM. The three phases are (1) skin segmentation based on a selected method and skin color space, (2) morphological-based image filtering for discarding non-face pixels, and (3) object analysis (for discarding non-face objects) and extraction (retrieve graphical and descriptive object information). The first two phases are performed on the whole image. The third phase is a smoothing step where attention is focused on each object. The basic framework of our approach is presented in Figure 2.



**Figure 2** Basic framework of proposed algorithm.

#### 3.1 Skin Segmentation

Skin segmentation separates pixels representing skin from background pixels. Table 1 shows the rule used for segmenting skin. If the pixel intensity conforms to the rule, then the pixel is considered as a skin pixel.

The detection of skin pixel can be performed using an explicit skin definition or using skin criteria derived from training data. All processes are performed in the specified user-selectable color spaces. These color spaces are RGB, normalized RGB, HSV and YCbCr. The implementation of color spaces are performed in two ways: cascade (serial) or ensemble (merge). The skin pixel detection is performed as follows:

$$S = I(x, y) \cap \Omega(C, \Theta) \quad (1)$$

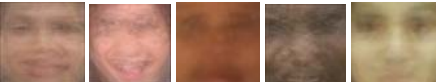
where  $S$  is decision that a pixel intensity in the specified location  $I(x, y)$  is in the criteria  $\Omega(C, \Theta)$  using skin segmentation method  $\Omega$  in the color space  $C$

with threshold value  $\Theta$ . Two methods can be employed for skin segmentation: (1) Piecewise Linear Decision Boundary (PLDB) and (2) Gaussian Mixture Models (GMM).

PLDB is a classification method that classifies two classes by using multiple decision rules (called decision boundary) so that the area is linearly separable in piece-wise fashion. This method is extended by utilizing color spaces. PLDB works in RGB color space. Because RGB representation does not separate luminance and chrominance components, it is extended by using chromaticity color spaces such as normalized RGB, HSV and YCbCr. The detail description of color spaces and their usage in detecting face were reported in [8], [9], [10], [11].

GMM is a parametric-based skin segmentation method. It assumes that skin distribution is concentrated in a specific area of image and has Gaussian characteristics. The model's parameter in this method is generated from training data, which are samples of skin images. The detail description of this method can be found in [12], [13], [14].

**Table 1** Skin segmentation decision table.

Methods	Basic Parameters	Decision Criteria
PLDB	RGB color space	$R > 95, G > 40, B > 20$ $\{\text{Max}(R,G,B) - \text{min}(R,G,B)\} < 15$ $ R - G  > 15, R > G, R > B$ $r/g > 1.185$
	RGB normalized color space	$(r*b)/(r+g+b)^2 > 0.107$ $(r*g)/(r+g+b)^2 > 0.112$
	HSV color space	$I > 40$ if $13 < S < 110, 0^\circ < H < 28^\circ$ and $332^\circ < H < 360^\circ$ if $13 < S < 75, 309^\circ < H < 331^\circ$
	YCbCr color space	$77 \leq Cb \leq 127$ and $133 \leq Cr \leq 173$
	GMM	Skin model 01 to skin model 05
GMM	Skin model 06 to skin model 10	

### 3.2 Minkowski Filtering Technique

Skin segmentation isolates face, non-face (e.g., hand, foot, etc.) and many other skin-like objects (i.e., those with similar skin color). There could be many orphan pixels that have no connectivity with other pixels and are concentrated in small region. Similarly, large pixel regions that do not conform to facial characteristics may also be found. In the image processing realm, this kind of regions can be considered as noise because it will give false positive detection result. To alleviate this problem, a mathematical-based morphological technique is utilized. The method employs a union-intersection set concept, also known as the Minkowski's addition and subtraction.

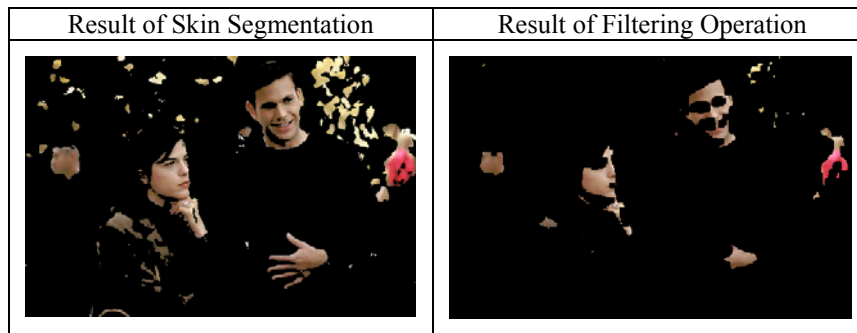
There are two basic Minkowski operators: dilation and erosion. Let  $A$  be a set of Euclidean coordinates that correspond to binary image. Let  $B$  be a structuring element that consists of small determination elements for erosion or dilation processes. Let  $B_x$  be a transformation process (movement, translation) of  $B$  for all image region.

Erosion  $A$  by  $B$  can be defined as a set of all points in  $x$  such that  $B_x$  is a subset of  $A$ .

$$\text{Erosion: } A - B = \bigcup_{b \in B} A_b = \{p \mid B_p \cap A\} \quad (2)$$

Dilation  $A$  by  $B$  can be defined as a set of all points  $x$  such that the intersection between  $B_x$  and  $A$  is not empty.

$$\text{Dilation: } A \oplus B = \bigcup_{b \in B} A_b = \{p \mid B_p \cap A \neq \emptyset\} \quad (3)$$



**Figure 3** Sample of Minkowski's filtering operation.

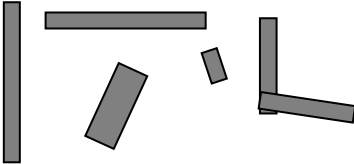
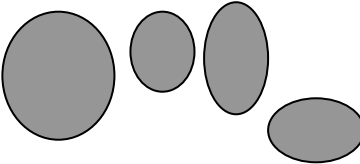


Basic morphological operators can be combined to obtain new operators. For example, the combination of erosion followed by dilation generates an opening operator. Meanwhile, dilation operation followed by erosion operation produces a closing operator. The use of morphological mathematics for image

enhancement was described in [15]. Figure 3 provides an example of morphological operation applied after skin segmentation process. As shown in the figure, the filtering operation removes orphan and non-face pixels.

### 3.3 Geometrical Object Shape Analysis

The analysis of geometrical object shape is performed on a specified object in the image. Each object is geometrically validated. Face is an ellipse-shape object in various size and position. An object that does not conform to geometrical criteria of face will be removed. Table 2 depicts non-face objects elimination process. In this example, non-ellipse objects such as finger, hand, foot and part of body will be deleted.

**Table 2** Sample of face and non-face geometrics criteria.

Non-face objects	Face or facial objects
	
	

Note: Black color in background is not part of face/non-face element

**Table 3** Eccentricity Determination Factors.

Shape factors	$f_0 = b/a$
Flattening	$f = (a-b)/a$
Linear eccentricity	$e = \sqrt{a^2 - b^2}$
Numerical eccentricity	$\varepsilon = e/a, \varepsilon' = e/b,$
Compactness	$C = 4\pi \cdot \text{area} / \text{perimeter}^2$
Roundness	$R = 4\pi \cdot \text{area} / (\text{convexperimeter})^2$
Elongation	$E = \text{width}_{\text{boundingbox}} / \text{height}_{\text{boundingbox}}$
Bending energy	$E_c = \frac{1}{L} \sum_{p=1}^L \kappa(p)^2$

Moment can be used for obtaining invariant value of an object [16], [14]. An object in digital image has at least three properties: size, position and orientation. Moment technique can extract these properties using several order moments. The zeroth order moment is used for obtaining the object size. The first order moment is used for extracting centroid and border information. The second-order moment can be used for orientation detection.

In our research, zeroth-order moment is applied after converting color image into binary image. Binary image is used in all moment processing because it is efficient and gives the best performance result. Table 4 describes the moment formula.

**Table 4** Moment calculation for digital image.

<b>Zeroth-order moment</b>	$m_{00} = \sum_x \sum_y f(x, y)$	$m_{10} = \sum_x \sum_y xf(x, y)$
	$m_{01} = \sum_x \sum_y yf(x, y)$	$m_{11} = \sum_x \sum_y xyf(x, y)$
<b>First-order moment</b>	$\bar{x} = \frac{m_{10}}{m_{00}}$	$\bar{y} = \frac{m_{01}}{m_{00}}$
<b>Second-order moment</b>	$\theta = \frac{1}{2} \tan^{-1} \frac{2\mu_{11}}{\mu_{20} - \mu_{02}}$	$\mu_{11} = m_{11} - \frac{m_{10}m_{01}}{m_{00}}$
	$\mu_{02} = m_{02} - \frac{m_{01}^2}{m_{00}}$	$\mu_{20} = m_{20} - \frac{m_{10}^2}{m_{00}}$

Vertical projection is part of integral projection. Integral projection can be used to represent the visual appearance of a certain kind of object under a relatively wide range of conditions, i.e., to model object classes. In this way, object analysis can be performed by fitting a test sample to the projection model. Projections of a binary image indicate the number of 1 pixels in each column, row, or diagonal in that image. We refer to them as horizontal, vertical, or diagonal projections, respectively. Although projections occupy much less memory than the image they were derived from, they still contain essential information about it. Figure 4 illustrates the vertical projection, which is focused on the skin color.





**Figure 4** Illustration of vertical projection.

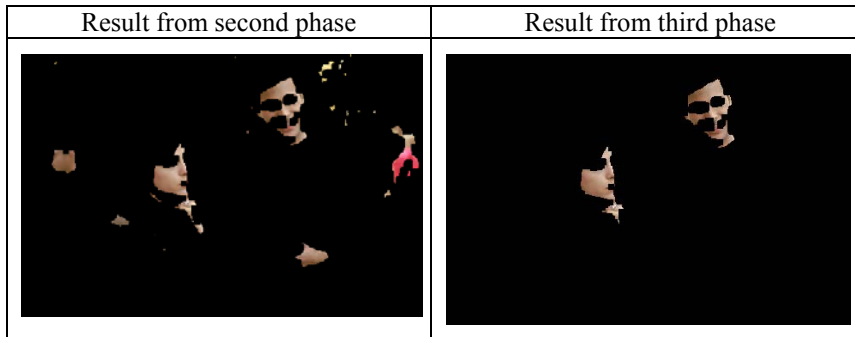
Vertical projection can be formulated as follows. Let  $I(x, y)$  be a pixel intensity and  $R(i)$  be a region in this image, i.e., a set of contiguous pixels in the domain of  $i$ . The vertical integral projections of  $R(i)$ , denoted by  $P_V(i)$ , is discrete and finite 1-D signals given by

$$P_{V(i)} : \{y_{\min}, \dots, y_{\max}\} \rightarrow \mathfrak{R}; P_{V(i)}(y) = |Ry(i)|^{-1} \sum_{x \in Ry(i)} i(x, y) \quad (4)$$

where  $y_{\min} = \min_{(x,y) \in R(i)} y$ ;  $y_{\max} = \max_{(x,y) \in R(i)} y$  and

$$Ry(i) = \{x / \forall x, (x, y) \in R(i)\}$$

Figure 5 shows the result of object-shape geometrical analysis step of our algorithm.



**Figure 5** Illustration of object shape and geometry analysis.

### 3.4 Information Extraction and Localization

As we see in Figure 5, objects in the image are cleaned and only face objects are available in the image. The next step is to extract information about the objects, including the number, size, location, bounding box and orientation of the face

objects. Based on this information, face objects can be localized, marked and cropped automatically in controllable situation. The individual face-object can then be used for other applications such as local enhancement of face objects, face motion detection, face feature detection, face recognition, face mapping, creating face print, avatar process, image filtering on the internet, to name a few.

Figure 6 shows the results of our proposed algorithm for detecting multi-faces objects in non-uniform, complex, real time scene.



**Figure 6** Localization and Information Extraction Result.

#### 4 Experiment and Performance Evaluation

Several experiments have been conducted to evaluate the performance of our approach. The data set used in the experiments contains images with single face object and multiple face objects. The data set does not contain image with non-standard face skin color or image whose background is similar to skin color. Most images were obtained from FERET database [17] and the Internet. Some of them were manually captured using Minolta S304 digital camera and Logitech Pro 4000 professional webcam. Table 5 describes the data set. All images in the data set possess the following characteristics: (i) complex, non-uniform background, (ii) face-object with various position, size and orientation, and (iii) object occlusion situation.

**Table 5** Description of data set.

	<b>Single Face</b>	<b>Multiple Faces</b>
Numbers of Image Files	7	33
Total Size	478KB	4.502KB
Image type	BMP, PNG, JPG	
Resolution	Vary from 208x369 up to 800x800, 24bpp	

The performance of algorithm is measured in terms of detection accuracy and detection speed. Accuracy can be defined as a statistical measure of how well a binary classification test correctly identifies or excludes a condition, as defined in the following:

$$A = \frac{\text{numbers of } TP + TN}{\text{numbers of } TP + FP + FN + TN} \quad (5)$$

where TP is True Positive, TN is True Negative, FN is False Negative, and FP is False Positive. The detection speed can be defined as the time required for a system to produce the detection results since it receives the input image. The experiment was run in a PC with Pentium IV 1.8 processor and 512MB memory.

Table 6 summarizes the experiment results. Based on these results, our algorithm can successfully detect all faces in single-face images. The detection speed varies depending on the image size. The detection accuracy in multi-face images is 91% in average. The detection speed in multi-face image is less than one second.

**Table 6** Performance Result if Face Detection Algorithm.

<b>Face</b>	<b>Number of Faces</b>	<b>Detection result</b>	<b>% accuracy</b>	<b>Time (second)</b>
Single	7	7	100	0.32
Many	220	199	91	0.99

## 5 Conclusion and Future Work

In this paper we have described our proposed algorithm for detecting faces in digital image. Unlike other similar algorithms, our algorithm is able to detect multiple faces in complex, non-uniform background at a reasonably fast detection speed. However, our algorithm will fail to detect a face if the image background is similar to skin color. It is also unable to detect a face painted with colors other than natural skin colors. These issues leave a lot of room for improvement.

Despite our preliminary success, the effectiveness of our algorithm remains an open research question. Further experiments that compare our method with other similar algorithms need to be conducted in order to address the issue.

## 6 Bibliographies

- [1] Yang, Ming-Hsuan., Kriegman, David J., & Ahuja, Narendra, *Detecting Faces in Images: A Survey*. *IEEE Transactions on PAMI*, **24**(1), January 2002.
- [2] Yang, G. & Huang, T.S., *Human Face Detection in Complex Background*, *Pattern Recognition*, **27**(1), pp. 53-63, 1994.
- [3] Yang, Ming-Hsuan, *Recent Advances in Face Detection*, International Conference on Patterns Recognition (ICPR), 2004.
- [4] Pantic, Maja and Rothkrantz, Leon J M., Automatic analysis of facial expressions: The state of the art, *IEEE Transactions on Patterns Analysis and Machine Intelligence (PAMI)*, **22**(12), pp. 1424-1445, 1996.
- [5] Sato, Atsuhshi & Imaoka, Hitoshi & Suzuki, Tetsuaki & Hosoi, Tohinori, *Advances in Face Detection and Recognition*, NEC Journal of Advance Technology, 2005.
- [6] Kotropoulos, C. & Pitas, I., *Rule-based Face Detection in Frontal Views*, Proc. International Conference on Accoustic, Speech and Signal Processing, **4**, pp. 2537-2540, 1997.
- [7] Zhao, Wen Yi, Chellappa, Rama, *Image-based Face Recognition: Issues and Methods*, Sarnoff Corporation and University of Maryland, 2004.
- [8] Hadi, Setiawan, Suwandi, Adang, Supriana, Iping & Wazdi, Farid, *Mathematical Model of Skin Color for Face Detection*, in proceedings International Conference on Applied Mathematics ICAM 2005, ITB, Bandung, Indonesia, 22-26 August 2005.
- [9] Hadi, Setiawan, et.al., *Skin Color Distribution Modeling for Pixel Classification in Chrominant Color Spaces*, National Journal Gematek STIKOM, March 2007 (in Bahasa Indonesia).
- [10] Hadi, Setiawan, Domesti, Rimba, Afifah, Wirda Noor, Susanti, Riska & Ismayadi, Isep, *Skin Color Information Utilization For Detecting Face In Digital Image*, Final Research Report of Program Hibah Kompetisi A2 Mathematics Department Padjadjaran University, 2006.
- [11] Hadi, Setiawan, Suwandi, Adang, Supriana, Iping & Wazdi, Farid, *Parametric Skin Distribution Modeling using Elliptical Boundary Model*, in proceedings of International Conference on Statistics and Mathematics (ICOMS) 2006, Bandung, 19-21 June 2006.
- [12] Hadi, Setiawan, *Image Segmentation for Extracting Facial Skin Information in Digital Image*, First Technical Report, Informatics STEI ITB, 2004 (in Bahasa Indonesia, published internally).

- [13] Hadi, Setiawan, Suwandi, Adang, Supriana, Iping & Wazdi, Farid, *Generating Skin Distribution Map of Face Images*, Seminar Nasional Aplikasi Teknologi Informasi SNATI 2006, Yogyakarta UII, 17 June 2006.
- [14] Tek, Cho Huak, Chin, Roland T., *On Image Analysis by the Methods of Moments*, IEEE Transactions on PAMI, **10**(4), July 1988.
- [15] Hadi, Setiawan, *Study, Development and Implementation of Minkowski Operator for Improving Object Detection Quality and Performance in Digital Image*, Second Technical Report, Informatics STEI ITB, 2005 (in Bahasa Indonesia, published internally).
- [16] Hadi, Setiawan, *Geometric-based of objects in digital image for face detection*, Third Technical Report, Informatics STEI ITB, 2006 (in Bahasa Indonesia, published internally).
- [17] NIST, *Standard Color FERET Facial Image Database*, National Institute of Standards and Technology, Image Group, Information Division, ITL, October 2003, in 2DVDs directly obtained from NIST.