

REVIEW

Open Access

# Multi-scale elastic graph matching for face detection

Yasuomi D Sato<sup>1,2,3\*</sup> and Yasutaka Kuriya<sup>1</sup>

## Abstract

We propose a multi-scale elastic graph matching (MS-EGM) algorithm for face detection, in which the conventional EGM is improved with two simple image processing techniques of the Gabor wavelet-based pyramid and the weak Gabor feature elimination. It is expected to solve difficulties of the real-time process in the conventional EGM. The Gabor wavelet-based pyramid effectively reduces not only the computational cost of the Gabor filtering but also the computational complexity of feature representation of a model face, preserving the facial information. The elimination of the weak Gabor feature extracted from an input image facilitates an accuracy of the Gabor feature similarity computations as unexpected. We then test that the MS-EGM can be capable of rapid face detection processing while achieving a high correct detection rate, comparable to the AdaBoost Haar-like (HL) feature cascade. We also show that the MS-EGM has strong robustness to the image of a face occluded with sunglasses and scarfs because of topologically preserved feature representations.

**Keywords:** Face detection; Elastic graph matching; Gabor wavelet-based pyramid; Weak Gabor response elimination

## 1 Review

### 1.1 Introduction

Face detection is of importance in various research fields for computer vision technology. For example, in a security system, it is an important preprocessing for identifying or tracking moving persons on an image, taken with a surveillance camera in public places. It may also often be used for acquiring information about a person in order to make more sophisticated dialogs between persons and life support robots [1].

As referred to [2,3], many different types of face detection algorithms have so far been suggested and developed progressively toward practical applications in industry. The algorithms are often classified into some approaches such as the knowledge-based [4], the feature-based [5,6], and so forth. In particular, for the feature-based approach, finding faces using random labeled graph matching has been well known [6]. Finding correct geometric arrangement of features for a face, the position of the face is

effectively identified with random labeled graph matching. However, the deficit of the random labeled graph matching is extremely difficult detection of facial features in a complex background.

Such face detection algorithms have been comparatively tested thus far [7]. One of them, remarkably exploited in recent computer vision, is a cascade of boosted classifiers with Haar-like (HL) features, proposed by Viola and Jones [8,9]. In the Viola-Jones algorithm, a set of weak classifiers is used. A number of weak classifiers are singled out and then organized into a cascade. Viola and Jones then succeeded to realize rapid face detection. Nevertheless, there still remain challenging problems to solve, one of which is the low detection performance problem when lower resolution images [10] or occluded facial images are used.

In face recognition processing, not only a face detector as the preprocess of the face recognition but also the face recognition itself is strongly significant. Many different algorithms for face recognition were developed, some of which must be well known as face recognition literature, called the elastic graph matching (EGM) [11-13] or the elastic bunch graph matching (EBGM) [14-16]. Their essential concept for effective facial identification is the so-called dynamic link architecture [17,18] where Gabor feature detectors establish dynamic projections of

\*Correspondence: sato-y@brain.kyutech.ac.jp

<sup>1</sup>Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4, Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan

<sup>2</sup>Frankfurt Institute for Advanced Studies (FIAS), University of Frankfurt, Ruth-Moufang-Str.1, Frankfurt am Main 60438, Germany

Full list of author information is available at the end of the article

an undistorted graph to the elastic one. In the algorithm, there seem to exist several demerits as follows: (1) topological feature-based representations' complexity and (2) the still difficult real-time process.

For the former, facial images are conventionally expressed with full Gabor feature representations encompassing many different orientations and many different spatial frequencies of the Gabor wavelet kernel [19]. This can easily be expected to cause consuming time increments and high computational complexity, which make the real-time process difficult. Many researchers have been trying to solve such problems by using complex wavelets [19], weighted sub-Gabor [20] wavelets, and simplified Gabor wavelets [21]. However, most of the problems have not been solved. It is still questionable what effective and efficient Gabor feature representations are.

Wolfrum et al. suggested a recurrent network model for processing of both detection and identification of a face on an input image [22]. The network model may be very interesting, whose technical improvements are furthermore requested toward practical applications. The Viola-Jones algorithm for face detection is integrated with the EGM algorithm for convenience [23,24]. Such an integrated system may be good. However, it is better that a face detection algorithm is established in the same concept for the face recognition algorithm of the EGM.

For this purpose, the current urgent task is to progressively develop an EGM-based face detection algorithm, which has rapid and high correct detection abilities, comparable to the AdaBoost of the HL cascade type. The final goal is to realize the real-time EGM. The achievement of the real-time EGM has so far been regarded to be extremely difficult. In this work, we discuss how the real-time process can be achieved, by using the simple image processing technologies of the Gabor pyramid and weak Gabor feature elimination.

In Section 1.2, we suggest a face detection algorithm based on combinations of the EGM and image pyramids. The overview of the face detection algorithm is shown in Figure 1. We may call it the multi-scale EGM (MS-EGM). In the MS-EGM, effective and efficient Gabor feature representations are of essence when the conventional EGM is improved with a Gabor wavelet-based pyramid [25,26]. This is mainly based on removal of the unnecessarily weak Gabor feature as well as on reduction of the complexity of the Gabor feature representation, preserving a critical amount of the facial information. Both the removal and reduction entirely simplify the system structure for face detection. In particular, the elimination of the unnecessarily weak Gabor feature is helpful for finding the desired location on the input (I) image, which is highly similar to model (M) facial representation. Thereafter, the EGM effectively detects the optimal location taking the highest similarity to the feature representation for the M face. It is

noticed that the elimination has an additional effect on the reduction of computational cost for face detection. Naturally, the same is for Gabor pyramid. Thus, the rapid face detection is achieved, comparable to the AdaBoost type of the HL features' cascade included by default in OpenCV 2.0: *haarcascade\_frontalface\_alt.xml* trained by Lienhart and Maydt [27].

In Section 1.3, we test the capability of face detection for the MS-EGM, in comparison with the one for the Viola-Jones algorithm. At least, four datasets of BioID [28], FERET [29], Caltech Faces [30], and AR Faces [31] are employed. Finally, discussion and conclusion will be given in Sections 1.4 and 2, respectively.

## 1.2 System design for face detection

A whole configuration for face detection in the MS-EGM is shown in Figure 2. The system basically consists of the preprocess for down-sampling and Gabor wavelet transform (GWT) and the main process for face detection with graph matching. The details about the preprocess and main process in the MS-EGM algorithm will be explained below.

### 1.2.1 Weak Gabor feature elimination

Let us explain preprocess for image I. Image I is rescaled with a size of  $60 \times 60$  pixels because the rescaled image with less computational cost of Gabor filters is more helpful than the original size. It is then convolved with a family of Gabor functions  $\psi_r(\mathbf{x}_1)$ :

$$\hat{J}_r^1(\mathbf{x}_1) = \int F(\mathbf{x}_1 - \mathbf{x}'_1) \psi_r(\mathbf{x}_1 - \mathbf{x}'_1) d^2 \mathbf{x}'_1, \quad (1)$$

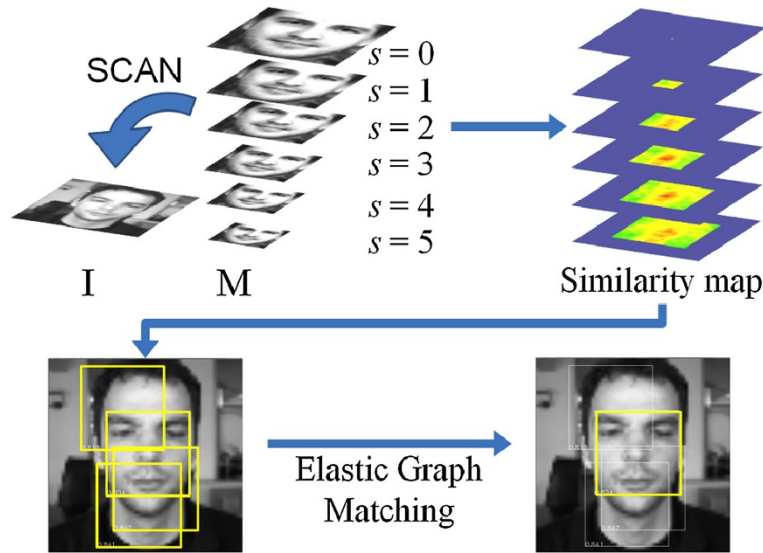
$$\psi_r(\mathbf{x}) = \frac{|\mathbf{k}_r|^2}{\sigma^2} \exp\left(-\frac{|\mathbf{k}_r|^2 |\mathbf{x}|^2}{2\sigma^2}\right) \times \left[ \exp(i\mathbf{k}_r \cdot \mathbf{x}) - \exp\left(-\frac{\sigma^2}{2}\right) \right], \quad (2)$$

where  $F(\mathbf{x})$  represents a gray scale value at pixel  $\mathbf{x}$  on I.  $r (= 0, \dots, 7)$  is an orientation parameter.  $\sigma = 2\pi$ . The wave number vector  $\mathbf{k}_r$  can be expressed as

$$\mathbf{k}_r = \begin{pmatrix} k_{r,x} \\ k_{r,y} \end{pmatrix} = \frac{\pi}{2} \begin{pmatrix} \cos \phi_r \\ \sin \phi_r \end{pmatrix}, \quad \phi_r = \frac{\pi}{8} r. \quad (3)$$

Both real and imaginary parts of the Gabor kernel are shown in Figure 3. In general, we will have to extrapolate the values of some non-existing pixels during the filter processing of the leftmost pixels of the image. In this work, the constant border extrapolation method was employed in OpenCV, letting us assume that all the non-existing pixels are zeros. The magnitude is used here as one of the feature values, that is,

$$J_r^1(\mathbf{x}_1) = \left| \hat{J}_r^1(\mathbf{x}_1) \right|. \quad (4)$$



**Figure 1** An overview of the face detection process. An input ( $I$ ) image is resized to  $60 \times 60$  pixels while the model ( $M$ ) image is down-sampled from  $60 \times 60$  pixels with  $a_0^{2s}$  where  $s = 0, \dots, 5$  and  $0 < a_0 < 1$ . The Gabor features are extracted from these images so that each filtered image  $M_s$ , on which a square graph is set up, scans over image  $I$  to calculate similarities between them. A similarity map pyramid is then formed to select several candidate locations that can be expected as highly similar to the  $M$  feature representation. Finally, the EGM for each candidate is pursued to single out the most likely location.

Here  $J_r^I(\mathbf{x}_1)$  is subject to a threshold condition for elimination of the weak response:

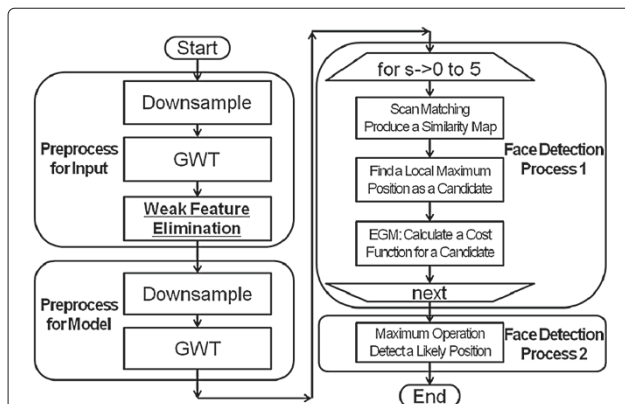
$$J_r^I(\mathbf{x}_1) = \begin{cases} 0 & \text{if } J_r^I(\mathbf{x}_1) < \Theta, \\ J_r^I(\mathbf{x}_1) & \text{else} \end{cases} \quad (5)$$

Figure 4 demonstrates Gabor features before and after the weak response elimination. The Gabor features seem to have no clear difference. However, the weak response

elimination importantly influences to face detection performance. The reasonable explanation will be given in the next section.

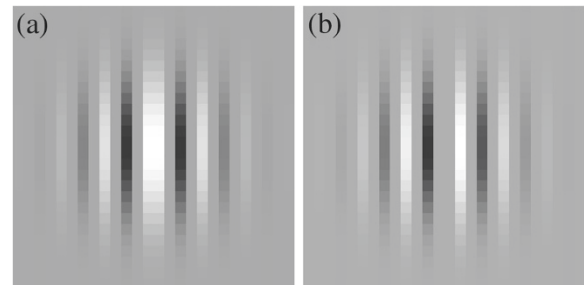
### 1.2.2 Gabor wavelet-based pyramid

In the next preprocess for the  $M$  facial image (Figure 2), we employ an image of the average face of German men created by a face generator [32]. The reason why the average face image is employed will be discussed in Section 1.4. The original size of image  $M$  is  $A_0 = 60 \times 60$  pixels (Figure 5). The image is down-sampled with  $a_0^s$  ( $s = 0, \dots, 5$ ), called the  $M_s$ .  $a_0 = 0.85$ .  $A_s = 60 \times 60 \times a_0^s \times a_0^s$ . The image  $M_s$  is convoluted with a family of Gabor functions, analogous to the preprocess for image  $I$ . This image

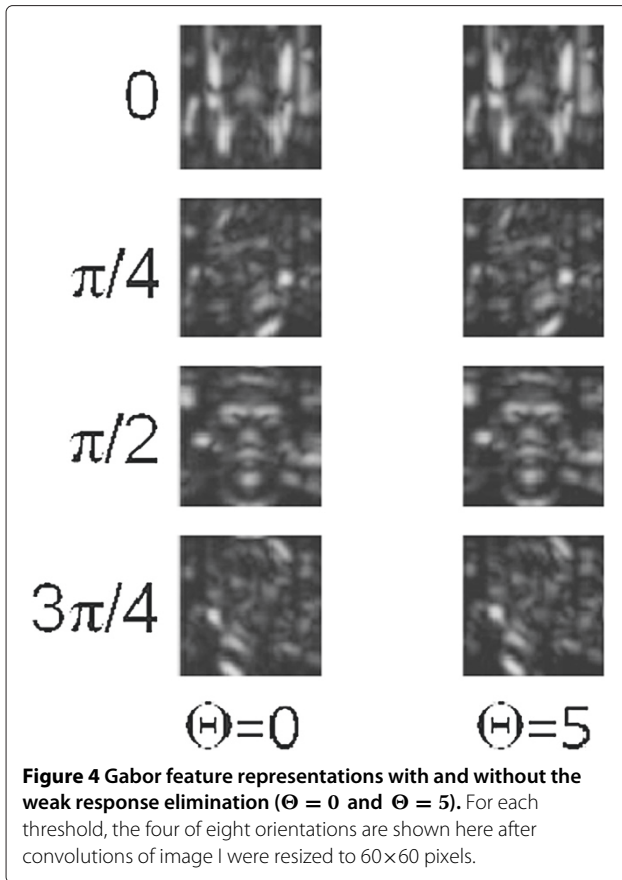


**Figure 2** A flow chart of the MS-EGM for face detection.

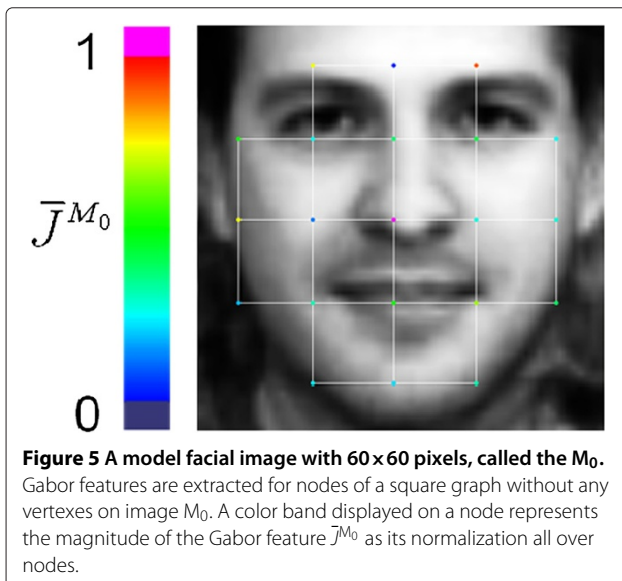
Preprocesses for images  $I$  and  $M$  are done with down-sampling and GWT. In the main process, an undistorted graph for each scale  $s$  is scanned over the whole  $I$  image for finding several candidate positions. The elastic graph is matched for detecting the most likely position in the candidates. The detailed procedure of this method will be described in the following section.



**Figure 3** Gabor kernels. (a) Real and (b) imaginary parts. The size of the Gaussian window is  $25 \times 25$  pixels.



processing, which has been widely known, is called as the Gabor wavelet-based pyramid (GWP). The Gabor feature is thus extracted, which takes one vector form consisting of eight orientation components  $J_r^{M_s}(\mathbf{x})$ , where  $r = 0, \dots, 7$ .



A square graph of  $(n \times n - 4)$  nodes without any vertexes is set on each resolution image  $M_s$ , letting the position of each node be  $\mathbf{p}_s$ . Here  $n$  ( $=5$ ) is the number of full nodes on a row and a column of the square graph. The graph is depicted in Figure 5. The reason why we removed these vertexes is that the MS-EGM has a risk that a correct detection rate declines with unnecessary Gabor feature for the vertex.

Let us briefly discuss the advantages of the use of the GWP, in comparison with the Gaussian pyramid employed in scale invariant feature transform (SIFT) [33,34], because there are two common points between the utilities of the GWP and Gaussian pyramid. One is the reduction of computational cost for the filters with sub-sampling images while the other is scale invariance.

The remarkable difference is the feature representation for a face. Because there is no orientation parameter on the Gaussian feature representation, the SIFT must be incorporated with additional representations such as histogram-of-oriented gradient (HoG) features [35,36]. Such duplicate feature representations for an image give rise to extremely high computational complexity and cost, although the SIFT with the HoG feature may be a logically sophisticated algorithm.

The multi-scaled Gabor feature is more abundant in facial representations, compared to the Gaussian feature. This indicates that such abundant facial representations on the Gabor feature can easily identify a location of the face on image I, which takes the highest similarity to feature representations of image M. This is shown in the next section, by creating a similarity map of image M for each size onto image I. One can see that higher similarity areas are predicted to correspond to locations of the face meanwhile the low similarities are regarded as the background. In any case, we herein address that an additional merit in the Gabor wavelet-based pyramid is the less computational complexity for an image information preserved in feature representations.

### 1.2.3 Face detection process

We shall proceed to the main process for face detection in Figure 2. The main process is basically carried out by EGM with a cost function  $E_s$  consisting of the similarity term  $e_s^s(\mathbf{x}_1)$  and the elastic term  $e_s^d(\mathbf{x}_1)$

$$E_s(\mathbf{x}_1) = e_s^s(\mathbf{x}_1) - \lambda_d e_s^d(\mathbf{x}_1), \quad (6)$$

$$e_s^s(\mathbf{x}_1) = \frac{1}{|G_s|} \sum_{\mathbf{p}_s \in G_s} \frac{\sum_r J_r^1(\mathbf{x}_1) \cdot J_r^{M_s}(\mathbf{p}_s)}{\sqrt{\sum_r (J_r^1(\mathbf{x}_1))^2 \sum_r (J_r^{M_s}(\mathbf{p}_s))^2}}, \quad (7)$$

$$e_s^d(\mathbf{x}_1) = \frac{1}{|G_s|} \sum_{\mathbf{p}_s \in G_s} \frac{\left| \sum_{\mathbf{p}'_s \in G'_s} D_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s} - \sum_{\mathbf{p}'_s \in G'_s} D_{\mathbf{x}_1, \mathbf{p}'_s}^I \right|}{\left| \sum_{\mathbf{p}'_s \in G'_s} D_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s} \right|} + 1 - \frac{1}{|G_s|} \sum_{\mathbf{p}_s \in G_s} \frac{\left| \sum_{\mathbf{p}'_s \in G'_s} B_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s} \cdot \sum_{\mathbf{p}'_s \in G'_s} B_{\mathbf{x}_1, \mathbf{p}'_s}^I \right|}{\left| \sum_{\mathbf{p}'_s \in G'_s} B_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s} \right| \cdot \left| \sum_{\mathbf{p}'_s \in G'_s} B_{\mathbf{x}_1, \mathbf{p}'_s}^I \right|}, \quad (8)$$

where  $e_s^d(\mathbf{x}_1)$  represents summation of the similarities of the Gabor feature at  $\mathbf{x}_1$  on image I to the one at node  $\mathbf{p}_s$ . In Equation 7, let us assume that the similarity is 0 when the norm equals 0 with the weak Gabor response elimination.  $G_s$  is a set of nodes on the graph.  $e_s^d(\mathbf{x}_1)$  represents the elasticity of the graph on image I.  $\lambda_d$  is a constant parameter for the graph elasticity.  $\lambda_d = 0.05$ , except for obtaining a similarity map  $E_s(\mathbf{x}_1)$  when  $\lambda_d = 0$ .  $G'_s$  is a set of nearest neighbor nodes  $\mathbf{p}'_s$  for  $\mathbf{p}_s$ .  $D_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s}$  and  $D_{\mathbf{x}_1, \mathbf{p}'_s}^I$  are the Euclidean distance between nodes  $\mathbf{p}_s$  (or  $\mathbf{x}_1$ ) and  $\mathbf{p}'_s$  on the graph of image  $M_s$  (or I).  $A_{\mathbf{p}_s, \mathbf{p}'_s}^{M_s}$  and  $A_{\mathbf{x}_1, \mathbf{p}'_s}^I$  take one vector form consisting of four elements. Each element is an angle between two nearest neighbors on each quadrant, centered at  $\mathbf{p}_s$ .

This face detection process is also split into two sub-processes: The first sub-process is undistorted graph matching with  $\lambda_d = 0$ , in which each node  $\mathbf{p}_s$  links to full pixels on I to produce a similarity map of  $\mathbf{J}^I$  to  $\mathbf{J}^{M_s}$  at each scale  $s$ . The local maximum positions on the map are then picked up as candidate locations of a face on I, whose feature representations closely resemble the  $M_s$  feature representations. The second is the real EGM with a finite  $\lambda_d$  to find the position taking the highest similarity in the candidates and, at the same time, to choose the most likely relative size of I to M.

**Select candidate positions** In the first sub-process, the upper left of image  $M_s$ , on which the graph is being set up, is firstly fitted to the one of I.  $M_s$  is repeatedly scanned to the right for each row on I. Then, the scan ends when the lower right of  $M_s$  arrives at the one of I (Figure 6a). As a result of the scan, we can observe a similarity map at each size of M image (see Figure 6b). Figure 7a shows one of the similarity maps of  $e_s^d(\mathbf{x}_1)$  obtained by scanning  $M_4$  over image I of Figure 6b.

We select some candidates that can be expected as a face position. Let the candidate position  $\mathbf{x}_1^{c_s}$  be defined as the center of the corresponding area with size  $A_s$  and label  $c_s$ . The candidate position is the local maximum on the similarity map, satisfied by a gradient condition when all differences of the  $\mathbf{x}_1^{c_s}$  to its nearest neighbors  $\mathbf{x}_1^{ne}$  take positive:

$$e_s^d(\mathbf{x}_1^{c_s}) - e_s^d(\mathbf{x}_1^{ne}) > 0. \quad (9)$$

The candidates  $c_s$  are furthermore confined to two or three  $c'_s$  at each scale  $s$  under an overlap condition. The overlap condition is defined as follows: Let the candidate positions  $\mathbf{x}_1^1$  and  $\mathbf{x}_1^2$  be close to each other. Let the size of an area  $O_1$  overlapping  $\mathbf{x}_1^1$  and  $\mathbf{x}_1^2$  be  $A_{O_1}$  (Figure 8). If  $[A_{O_1}/A_s] > 0.5$  and  $e_s^d(\mathbf{x}_1^1) - e_s^d(\mathbf{x}_1^2) > 0$ , the candidate  $\mathbf{x}_1^2$  should be ruled out. The selected candidate pixel is depicted with a black square on the similarity map as shown in Figure 7a.

We herein notice that the candidate confinement is naturally done for the sake of achievements of the rapid face detection processing. One can see that computational cost for the following EGM is increased with the increment of the candidate number.

**Find the most likely position** In the second sub-process, the EGM for each candidate position  $\mathbf{x}_1^{c'_s}$  is pursued for obtaining the maximum value of the cost function  $E_s^{c'_s}$ . Each node on image I, which corresponds to node  $\mathbf{p}_s^{c'_s}$  of the square graph  $M_s$ , surveys an optimal pixel  $\mathbf{p}_s^I$  taking a maximum of the cost function  $E_s^{c'_s}$  within a search region  $R$ :

$$\mathbf{p}_s^I = \max_{\mathbf{x}_1 \in R} \{E_s^{c'_s}(\mathbf{x}_1)\}, \quad (10)$$

where  $R$  is a set of pixels that can be picked up in a square with the size of  $(2q+1) \times (2q+1)$ , centered at the pixel corresponding to  $\mathbf{p}_s^{c'_s}$ .  $q = 4$ . Figure 9 shows achievement of the distorted graph on one candidate extracted from image I.

Finally, the optimal pixel  $\mathbf{x}_1^m$  is singled out with the maximum operation of all candidates  $c'_s$ , which must be the most similar to feature representation for the M face:

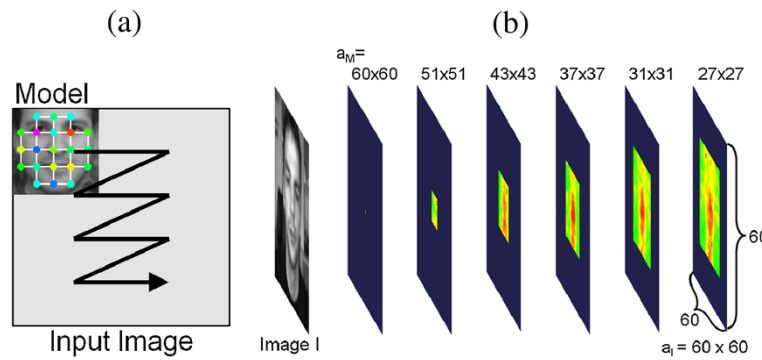
$$\mathbf{x}_1^m = \max_{\mathbf{x}_1^{c'_s} \in C} \{E_s^{c'_s}(\mathbf{x}_1^{c'_s})\}. \quad (11)$$

Here let  $C$  be a set of candidates  $c'_s$ .

### 1.3 Face detection ability tests

We test a face detection ability of the MS-EGM, using four different databases of BioID [28], FERET [29], Caltech Faces [30], and AR Faces [31], compared to an AdaBoost type of the HL cascade face detector mentioned in Section 1.1. Both face detectors can process on the following computer configuration: Intel Core i5-2410M Processor (2.30 GHz  $\times 2$ , TurboBoost 2.90 GHz, hyper-threading, TDP 35W), RAM 4 GB, Windows 7 Home Premium SP1 64 bit.

We assume that correct detection can be defined when the eyes are located in three fifths from the top of the resulting square area (as shown in the silhouette part of Figure 10a) whereas the incorrect detection



**Figure 6 A scan scheme and a similarity map pyramid. (a)** A scan scheme for each size of the M image in calculations of similarity maps. **(b)** A similarity map pyramid. Each similarity map is calculated by scanning each graph on M with  $60 \times 60$ ,  $51 \times 51$ ,  $43 \times 43$ ,  $37 \times 37$ ,  $31 \times 31$ , and  $27 \times 27$  (pixels) on I.

means that the eyes' position is out of the upper area (Figure 10b).

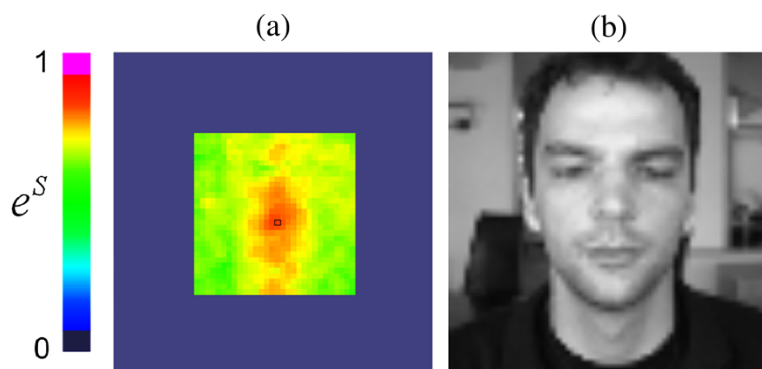
Essential mechanisms on ascending face detection performance will be fundamentally explained, using the BioID database. With the databases of FERET and Caltech Faces, we will verify that the MS-EGM face detector has a detection ability, comparable to the AdaBoost HL cascade. In the test using Caltech Faces, robustness for occluded face images will be indicated. Using the AR Faces database containing many occluded face images, it will be shown that the MS-EGM must be predominant over the AdaBoost HL cascade, in terms of occlusions. The final face detection performance result is shown in Table 1.

We notice that we employ only one competing method of the AdaBoost type of the HL cascade. This is because there remain a lot of improvements such as Gabor feature extraction with fast Fourier transform (FFT). Thus, we decided that after improving more the MS-EGM, we should test the detection ability, using the other databases that were not employed in this work.

### 1.3.1 Weak Gabor response elimination

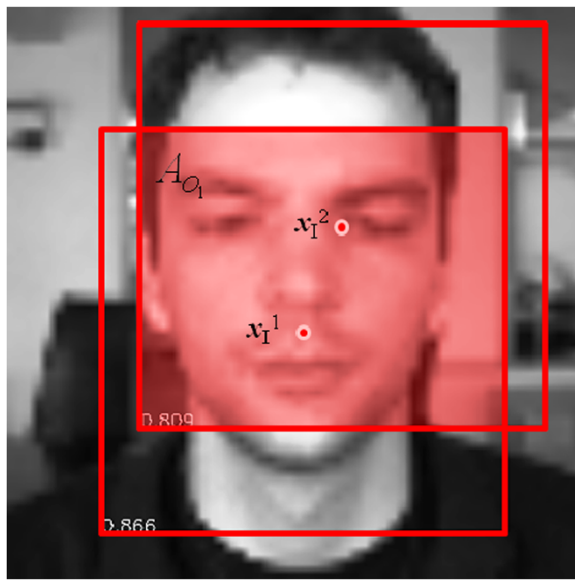
We see how face detection performance varies with changes of the threshold  $\Theta$ . When  $\Theta = 0$  without removals of any infinitesimal convolutions, we have obtained the detection rate, 96.8%. For  $\Theta = 2.5$ , the correct detection rate slightly increases to 97.5%. Nevertheless, the Gabor filters seem to be unaffected by designed operations of the threshold Figure 3. At  $\Theta = 5$ , the detection rate rises approximately up to 97.6%. However, it starts descending when the threshold further increases (see Table 2).

We will explain the reason why a face detection performance increases with eliminations of the weak components in the Gabor feature. Firstly, let us consider the case for  $\Theta = 0$ . We can find several pixels in the unwanted area on image I (Figure 11a), which take the relatively high similarities to the M Gabor features (Figure 11c). Next, letting the threshold  $\Theta = 5$  be given, some of the elements exhibiting weak responses in I Gabor features should be reset to 0. Correspondingly, the relatively high similarities are replaced to the lower values (Figure 11d) such that



**Figure 7 The results for one of the similarity maps. (a)** Similarities  $e_s^4$  are calculated by scanning a model graph  $M_4$  all over image I. A color band represents values of  $e_s^4$ . **(b)** The I image of a face with a background is resized to  $60 \times 60$  pixels.



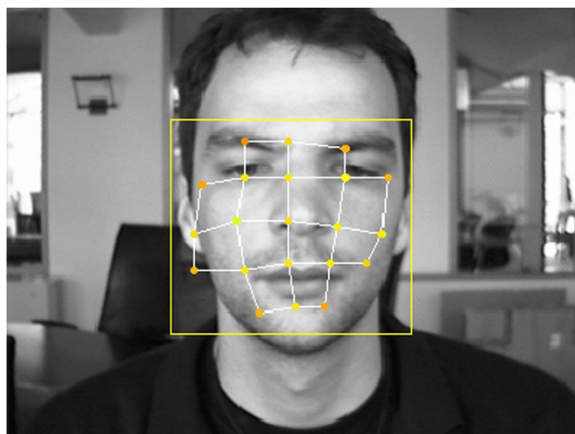


**Figure 8 An overlap condition to determine the candidate position.** The silhouette part with a size of  $A_{O_1}$  is the one overlapping two pre-candidate areas centered at  $x_I^1$  and  $x_I^2$ . Numerals on the bottom left are values of the similarity.

a desired pixel can be easily found on image  $I$  as shown in Figure 11b. This is also supposed to be important to accurately find a location of the face on image  $I$ . Finally, we mention an additional advantage of elimination of the weak Gabor response. In Table 2, consuming times in the case with setting of the threshold are reduced, compared to the case without setting.

### 1.3.2 Requirement of EGM

We may suppose that the MS-EGM does not need the EGM process (which means that the most likely position



**Figure 9 Achievement of face detection.** The EGM is computed after finding one candidate position, which is expected to be like the model feature representation.

is specified only by taking maximum in the candidates). This is, however, apparently false because the detection rate decreases from 97.6% to 96.0% (see Table 3). It has been shown that the local maximum of the similarity map is just a signpost to find the optimal position meanwhile the EGM is one of the necessary factors for detecting the most likely position.

### 1.3.3 Occlusion

We test detection abilities with other databases of the FERET and Caltech Faces. For the FERET, we have found that our system is almost comparable to the AdaBoost HL cascade. In the Caltech Faces, the AdaBoost HL cascade shows higher detection performance than our MS-EGM, which wins against the AdaBoost HL cascade, in terms of process speed. In the Caltech Faces, there are also some occluded face images. The AdaBoost HL cascade could not correctly detect the occluded faces whereas our MS-EGM did it very well.

Then, we tested the detection ability with the use of the AR Faces database. The AR Faces database was made by Computer Vision Center at Universitat Autònoma de Barcelona (UAB). It contains the images of 126 people (70 men and 56 women), where there are frontal faces with different facial expressions, illumination conditions, and occlusions (sunglasses and scarfs).

In Table 1, the AdaBoost HL cascade obtained 90% while our MS-EGM achieved a higher performance, 97%. For the average consuming time with the AR Faces database, the HL cascade costs less consuming time than our MS-EGM. The reason that the HL cascade can cost less consuming time is that the background of the occluded images is blank.

We consider why our MS-EGM can get a higher correct face detection rate than the AdaBoost HL cascade. In the AdaBoost HL cascade, incorrect detection mostly occurs with images of a face occluded with sunglasses or scarfs. This is due to that any suitable HL features, selected by the AdaBoost within a search window, can never fit the sunglasses nor the scarfs. Meanwhile in the MS-EGM, the graph is distorted on the sunglasses or the scarfs, with a unique property of the EGM, 'topological preservation.' However, the topologically distorted feature representation can still be recognized as the feature representation for a model face for achieving face detection, as shown in Figure 12.

### 1.4 Discussion

In this work, we have proposed the MS-EGM that should be worthy of comparison to the detection ability of the AdaBoost HL cascade. However, we still have to improve more the MS-EGM to integrate multi-face detection because the AdaBoost HL cascade could detect multi-objects. Also, the MS-EGM may still be requested for



**Figure 10 Definition of face detection.** Correct detection **(a)** and incorrect detection **(b)** show that the eyes are in and out of the three fifths from the top, respectively.

higher performance for the sake of its effective works in the real world. So, it would be necessary for us to inspect the causes of false detection in the experiments or drawbacks in our system.

The false detection was done mostly by detecting backgrounds. One reason for background detections is the square graph-based feature representation. As shown in Figure 1, most of the grids are located around fiducial points on the face, like eyebrows, eyes, lips, and facial edge. Such square graph-based feature representations are significantly sensitive to linear edges in the backgrounds. Indeed, they take higher similarities to representations for linear edges, compared to the facials. The square graph-based feature representations also induce misrecognition in that that our system recognizes the round necks and lips as the lips and eyes, respectively. In order to solve such misrecognition problems, it is better to use a face graph, instead of a square graph. This is because face graph-based feature representations can be distinguished against the background and may reduce misrecognition.

The other cause for misdetection is that the size of the input face cannot be identified because it is unfitted to any sizes of down-sampling model faces. So, if another size of the model face is additionally prepared, our system can more easily detect the suitable size of the input face. It can thus expect to increase the detection rate. However, we have to be aware of the trade-off between computational

costs and the number of the model face size to be prepared because an increase of face size number causes to take a longer process time. Then, we will have to find the optimal parameter such as the number of sizes and so on.

In order to increase the correct detection rate, it would be beneficial to employ not only the eyes but also other fiducial points such as the nose. As mentioned above, our face detection system with the square graph tends to misrecognize the lip to be the eyes. This indicates that the obtained square for face detection tends to be positioned below the eyes. The correct detection can be defined again if a silhouette part in the obtained square involves both the eyes and nose or only the nose. Since, in this redefinition, the nose in the silhouette part compensates for incorrect detection even though the eyes are out of the silhouette part, it can be expected to reduce misrecognition and then further increase the correct detection rate.

In this work, we employed only one competing method of the AdaBoost type of the HL cascade. We suppose that it is interesting to compare our MS-EGM with other methods of face detection such as SIFT. In the SIFT, the main functionalities are thought as follows: (1) to find key points to create features invariant to scale and rotation and (2) to create HoG feature representations from the key points. Such functionalities are preprocesses for face/object detection. So, in order to compare the SIFT with our MS-EGM, we will have to more actively discuss the following: (1) how to make feature representations for

**Table 1 Comparative performance evaluation for face detection**

Database	Total image number	Image size (pixels)	MS-EGM rate	MS-EGM time	AdaBoost HL cascade rate	AdaBoost HL cascade times
BioID	1,521	384×286	97.6	37	96.2	36
FERET	2,963	256×384	99.3	67	99.8	62
Caltech Faces	450	896×592	97.8	52	99.1	178
AR Faces	3,262	128×128	98.0	55	90.4	7

A correct face detection rate (%) and the average consuming time (ms) per one image for the MS-EGM (with  $\Theta = 5$ ) and AdaBoost HL feature cascade, respectively. They are calculated with BioID, FERET, Caltech Faces, and AR Faces.



**Table 2 Face detection performance for different thresholds**

$\Theta$	Rate ( % )	Time (ms)
0	96.8	41
2.5	97.5	37
5	97.6	37
7.5	96.0	33
10	88.0	31

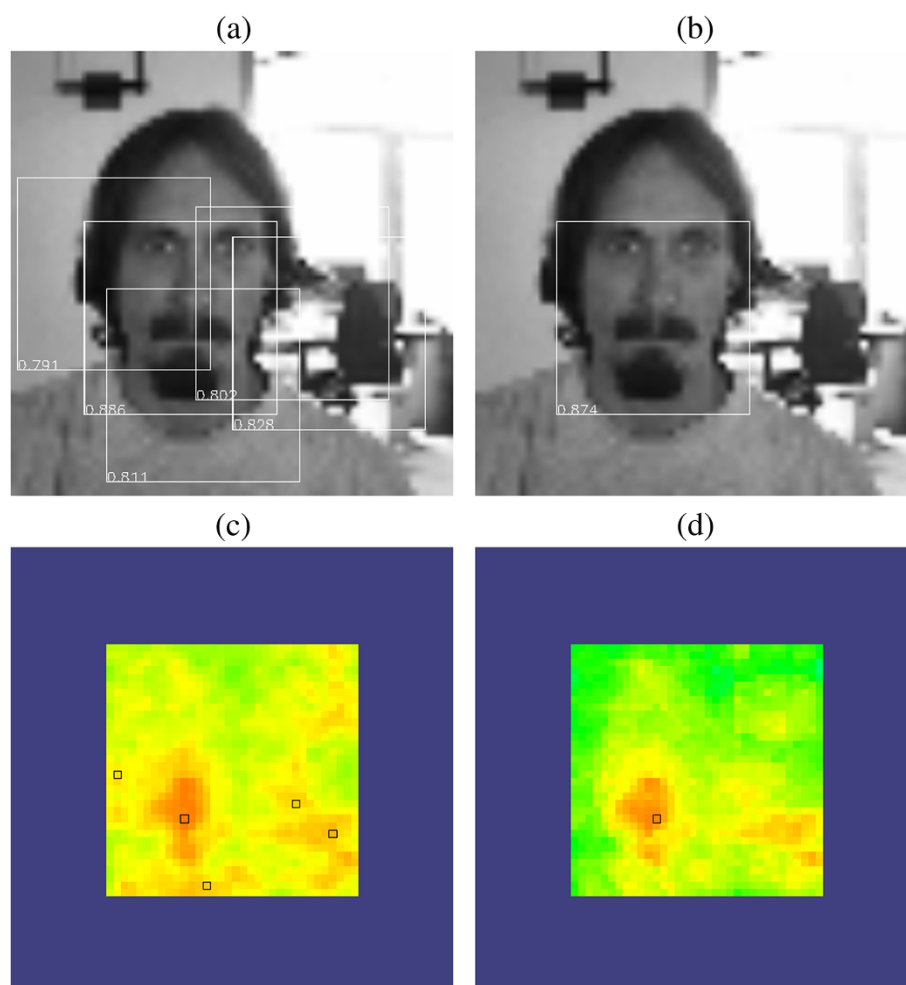
Changes of a correct face detection rate ( % ) and the average consuming time (ms).  
With threshold ( $\Theta$ ) operations in the MS-EGM, using BioID.

a model face and (2) how to establish the matching process for detection with SIFT-based feature representations. Such discussions will give us insights into improvements of our MS-EGM.

In addition, the model image in our MS-EGM algorithm was down-sampled, in a different way from the SIFT. This

can be predicted to cause more efficient scale invariance. When the optimal scale correspondence is to be considered in a Gabor pyramid in the MS-EGM, thus, it can expect to highly identify the size of a face on image I. At least, in the real-time face detection demonstration (not shown here), the MS-EGM is robust to illumination and distortion. In the future, introducing the rotation invariance [37,38] into our MS-EGM, we develop a more accurate face detection system that can easily detect the face when putting the head on one side. Also, it is interesting to use the face graph, instead of the square graph.

The deficits in our current algorithm are constraints to the down-sampled size of a model face and the incorrect detection on much lower resolution images. These are closely related to the Gabor kernel size of  $25 \times 25$  pixels employed in our system. In the current algorithm for convolution with the Gabor filter, if we use images



**Figure 11 Finding candidate regions on image I using image  $M_5$  (a, b) and similarity maps (c, d). (a, c)  $\Theta = 0$ . (b, d)  $\Theta = 5$ . In similarity maps, a black square shows the local maximum pixel. Numerals at the lower left inside the square are the local maximum values.**

**Table 3 Effects of EGM on correct face detection**

Database	With EGM	With EGM	Without EGM	Without EGM
	rate	time	rate	time
Bioid	97.6	37	96.1	12
Caltech Faces	97.1	52	96.6	21

Differences of a correct face detection rate (%) and the average consuming time (ms) per one image. With/without EGMs in the MS-EGM ( $\Theta = 5.0$ ), using Bioid and Caltech Faces.

smaller than  $25 \times 25$  pixels, we will need to implement extrapolation methods such as the replicated border and the constant one to cover additional pixels outside the images. However, even if such extrapolation methods are implemented into our current algorithm, it is still unclear whether or not our system can have better detection performance. This is because we do not ensure whether or not appropriately desired convoluted values are computed.

The MS-EGM face detector developed in this work saves a critical amount of information for a face but succeeds to remove the complexity of feature representation as possible as we can. The HoG features, which are, in recent years, used frequently for being incorporated with the SIFT, encompass all information around the local, which most probably contains the undesired information. Such information causes increments of computational cost. After all, computational costs increased by dealing with the undesired information have to be reduced.

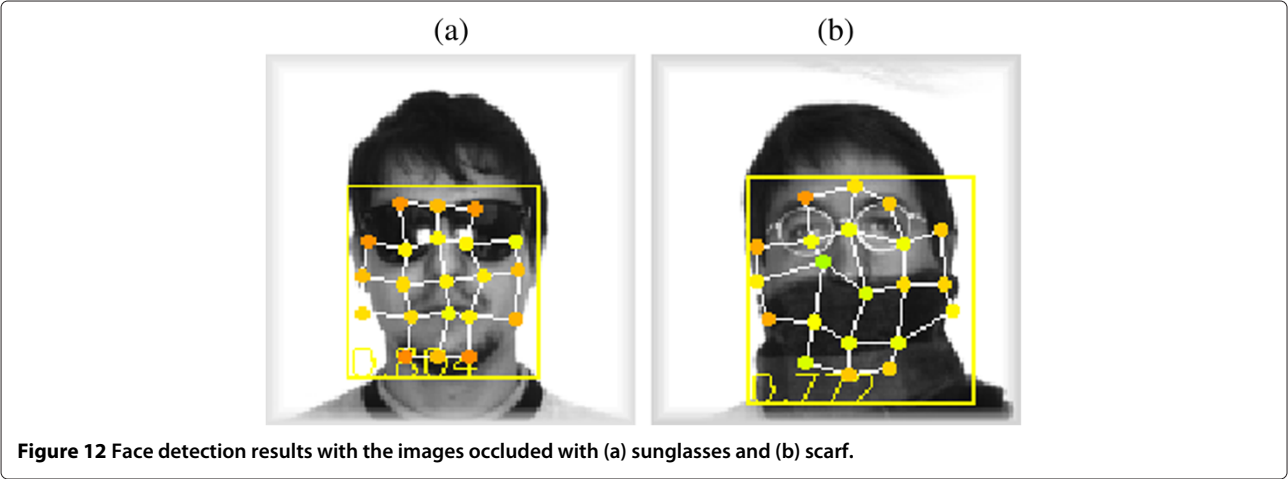
However, our system has realized rapid and high face detection performance without easygoing supports of the CPU power, taking into account the following points: improvements of Gabor feature representation, namely removals of complexity and noise on the representations. In particular, removing the weak Gabor responses has almost never been done as far as we know. It was shown that weak Gabor responses in high-resolution

satellite images were removed with Otsu’s thresholding method [39] in order to find edges of buildings [40]. Effects of such removals on detection or on recognition were not yet discussed. Meanwhile, we have found that it is a significantly effective technique for increasing the probability of correct detection. In the future, it may be more interesting to apply the weak Gabor feature elimination into the EGM or EBGM algorithm for visual object/face recognition.

It is also interesting for us to implement MS-EGM into a digital hardware circuit. This is because the EGM in our face detection algorithm is developed by modifying the EGM for face recognition improved into the digital hardware [41]. It can also expect that the EGM-specific integrated circuit will be developed for parallel processing for detection and recognition of a face. Such a system must be definitely compact and simplified, compared to the system integrated with several individual functions.

In this work, image I was resampled to  $60 \times 60$  pixels because of reduction of computational cost of the Gabor filter for the original size. The resampling to  $64 \times 64$  pixels would be better rather than that to  $60 \times 60$  pixels if we consider the extraction of GW features using the FFT. However, taking current computer performances into account, it can be expected that both cases of  $64 \times 64$  pixels and  $60 \times 60$  pixels obtain almost the same speed such that both cases can realize real-time performance of EGM.

Rather, we have to be aware that there is still one problem about a down-sampled scaling factor in the Gabor pyramid. One of the merits to use the Gabor pyramid is the optimal scale correspondence, which provides us the best trade-off between spatial resolution and frequency resolution. Nevertheless, we did not care of the optimal scale correspondence but only how the EGM is realized in the real-time process. As the next step, we will have to reconsider the correct scale for finding the best trade-off. This allows us to realize the additional functionality of



the scale invariance. We can establish the EGM system for face detection, independent of facial sizes of the input.

Finally, we discuss the reason to employ an image of the average face for German. We will have to be aware that the core target in our work is to develop the algorithm for parallel processing for detection and recognition of a face within the framework of the EBGM. In the EBGM, topologically the same face graphs are prepared for different identities in the M domain. Such different face graphs stored in the M domain recognize or identify the corresponding face on image I. Assuming that such an EBGM-based algorithm for face recognition is integrated with our algorithm for face detection in the future, an average face image must be better to use, rather than using several different face graphs or one identifiable face image. In this work, using an average face image for German men, we have found extra tasks, one of which is how racial and gender differences affect the face detection ability of our MS-EGM. This also implies that our MS-EGM algorithm proposed here is expected to further develop.

## 2 Conclusions

In this work, we have proposed the MS-EGM as a face detector, improving two following formulae in the conventional EGM: The first is to use a Gabor wavelet-based pyramid. This effectively reduces not only computational costs for Gabor filtering but also computational complexity for feature representation, preserving the image information about the model face. The second is elimination of the weak Gabor feature extracted for image I. This facilitates an accuracy of the Gabor feature similarity computations as unexpected. The MS-EGM can thus be capable of rapid face detection processing while achieving a high rate, comparable to the AdaBoost HL feature cascade. We have shown that the MS-EGM has strong robustness to the image of a face occluded with sunglasses and scarfs because of topologically preserved feature representations.

### Competing interests

The authors declare that they have no competing interests.

### Acknowledgements

The authors thank C. von der Malsburg at FIAS, C. Weber, S. Wermeter at the University of Hamburg, K. Horio, and H. Miyamoto at Kyushu Institute of Technology for the fruitful and active discussion. This work was partially supported by the Grant-in-Aid for Challenging Exploratory Research (to Y.D.S.) (no. 25540110) from MEXT.

### Author details

<sup>1</sup>Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4, Hibikino, Wakamatsu-ku, Kitakyushu 808-0196, Japan. <sup>2</sup>Frankfurt Institute for Advanced Studies (FIAS), University of Frankfurt, Ruth-Moufang-Str.1, Frankfurt am Main 60438, Germany. <sup>3</sup>Institute of Industrial Science, the University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan.

Received: 5 July 2013 Accepted: 1 November 2013  
Published: 21 November 2013

## References

1. Cruz, LE, Sucar, EF, Morales. Real-time face recognition for human-robot interaction, in *Proceedings of the 8th IEEE International Conference on Automatic Face and Gesture Recognition* (IEEE, Piscataway, 2008), pp. 1–6
2. E. Helmas, BK Low, Face detection: a survey. *Comput. Vis. Image Underst.* **83**(3), 236–274 (2001)
3. MH Yang, DJ Kriegman, N Ahuja, Detecting faces in images: a survey. *IEEE Trans. Pattern Anal. Mach. Intel.* **24**(1), 34–58 (2002)
4. G Yang, TS Huang, Human face detection in complex background. *Pattern Recognit.* **27**(1), 53–63 (1992)
5. R Kjeldsen, J Kender. Finding skin in color images, in *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition* (IEEE, Piscataway, 1996), pp. 312–317
6. TK Leung, MC Burl, P Perona. Finding faces in cluttered scenes using random labeled graph matching, in *Proceedings of the Fifth IEEE International Conference on Computer Vision* (IEEE, Piscataway, 1995), pp. 637–644
7. N Degtyarev, O Seredin. Comparative testing of face detection algorithms, in *Image and Signal Processing*, ed. by A Elmoataz, O Lezoray, F Nouboud, D Mammass, J Meunier. 4th International Conference, ICISP 2010, Trois-Rivières, QC, Canada, June 30–July 2, 2010. *Lecture Notes in Computer Science*, vol. 6134 (Springer Heidelberg, 2010), pp. 200–209
8. P Viola, MJ Jones. Rapid object detection using a boosted cascade of simple, in *Proceedings of the IEEE CVPR* (IEEE Piscataway, 2001)
9. P Viola, MJ Jones. Robust real-time face detection. *Int. J. Comput. Vis.* **57**(2), 137–154 (2004)
10. S Hayashi, O Hasegawa, Robust face detection for low-resolution images. *J. Adv. Comput. Intell. Inform.* **10**(1), 93–101 (2006)
11. M Lades, JC Vorbrueggen, J Buhmann, J Lange, C von der Malsburg, RP Wurtz, W Konen, Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* **42**(3), 300–311 (1993)
12. H Shin, S-D Kim, H-C Choi, Generalized elastic graph matching for face recognition. *Pattern Recognit. Lett.* **28**, 1077–1082 (2007)
13. YD Sato, Y Kuriya. Elastic graph matching on Gabor feature representation at low image resolution, in *Artificial Neural Networks and Machine Learning – ICANN 2012*, ed. by AEP Villa, W Duch, P Érdi, F Masulli, G Palm. 22nd International Conference on Artificial Neural Networks, Lausanne, Switzerland, September 11–14, 2012, Part I. *Lecture Notes in Computer Science*, vol. 7552 (Springer Berlin, 2012), pp. 387–394
14. L Wiskott, J-M Fellous, N Krüger, C von der Malsburg, Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 775–779 (1997)
15. D Monzo, A Albiol, J Sastre. HOG-EBGM vs. Gabor-EBGM, in *15th IEEE International Conference on Image Processing, 2008* (IEEE Piscataway, 2008), pp. 1636–1639
16. S Sarkar. Skin segmentation based elastic bunch graph matching for efficient multiple face recognition, in *Advances in Computer Science, Engineering & Applications*, ed. by DC Wyld, J Zizka, D Nagamalai. Second International Conference on Computer Science, Engineering and Applications (ICCSEA 2012), May 25–27, 2012, New Delhi, India, Volume 1. *Advances in Intelligent and Soft Computing*, vol. 166 (Springer Berlin, 2012), pp. 31–40
17. C von der Malsburg, *The correlation theory of brain function, internal report.* (Max-Planck-Institute für Biophysikalische Chemie, Göttingen, 1981)
18. C von der Malsburg. How are nervous structures organized?, in *Synergetics of the Brain*, ed. by E Basar, H Flohr, H Haken, A Mandell. *Proceedings of the International Symposium on Synergetics at Schloss Elmau, Bavaria, May 2–7, 1983.* Springer Series in Synergetics, vol. 23 (Springer Berlin, 1983), pp. 238–249
19. A Eleyan, H Ozkaramanli, H Demirel, Complex wavelet transform-based face recognition. *EURASIP J. Adv. Signal Process.* **2008**(185281), 13 (2008)
20. L Nanni, D Maio, Weighted sub-Gabor for face recognition. *Pattern Recognit. Lett.* **28**(4), 487–492 (2007)
21. W-P Choi, S-H Tse, K-W Wong, K-M Lam, Simplified Gabor wavelets for human face recognition. *Pattern Recognit.* **41**(3), 1186–1199 (2008)
22. P Wolfrum, C Wolff, J Lücke, C von der Malsburg, A recurrent dynamic model for correspondence-based face recognition. *J. Vis.* **8**, 1–18 (2008)
23. IR Khan, H Miyamoto, T Morie. Face and arm-posture recognition for secure human-machine interaction, in *IEEE International Conference on Systems, Man and Cybernetics (SMC2008)* (IEEE Piscataway, 2008), pp. 411–417

24. IR Khan, T Morie, H Miyamoto, M Shimizu, Y Kuriya. A prototype system for secure human-machine interaction based on face and gesture recognition, in *34th Annual Conference of the IEEE Industrial Electronics Society (IECON'08)* (IEEE Piscataway, 2008), pp. 1572–1577
25. YD Sato, J Jitsev, J Bornschein, D Pamplona, C Keck, C von der Malsburg. A Gabor wavelet pyramid-based object detection algorithm, in *Advances in Neural Networks – ISNN 2011*, ed. by D Liu, H Zhang, M Polycarpou, C Alippi, H He. 8th International Symposium on Neural Networks, ISNN 2011, Guilin, China, May 29–June 1, 2011, Part II. Lecture Notes in Computer Science, vol. 6676 (Springer Berlin, 2011), pp. 232–240
26. T Fukutomi, YD Sato, H Miyamoto. Multi-scale perception model for visual illusion on hybrid image, in *Proceedings of the Joint 6th International Conference on SCIS and 13th ISIS* (IEEE Piscataway, 2012), pp. 336–340
27. R Lienhart, J Maydt. An extended set of Haar-like features for rapid object detection, in *Proceedings of the International Conference on Image Processing* (IEEE Piscataway, 2002), pp. 900–903
28. O Jesorsky, K Kirchberg, R Frischholz. Robust face detection using the Hausdorff distance, in *Audio and Video-Based Person Authentication – AVBPA 2001*, ed. by J Bigun, F Smeraldi. Third International Conference on Audio- and Video-based Biometric Person Authentication, Halmstad, Sweden, 6–8 June 2001. Lecture Notes in Computer Science (Springer Berlin, 2001), pp. 90–95
29. PJ Phillips, H Moon, SA Rizvi, PJ Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Machine Intell.* **22**, 1090–1104 (2000)
30. Computational Vision Group, Caltech Faces (1991), <http://www.vision.caltech.edu/html-files/archive.html>.
31. AM Martinez, AC Kak. PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(2), 228–233 (2001)
32. The Face of Tomorrow, What is the face of London, Rio, Tokyo? What is the face of your city? (2013).
33. DG Lowe. Object recognition from local scale-invariant features. *Proc. Int. Conf. Comput. Vis.* **2**, 150–1157 (1999)
34. DG Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **60**(2), 91–110 (2004)
35. H Fujiyoshi. Gradient-based feature extraction, SIFT and HOG. Tech. Rep. IEICE. PRMU. **107**(206), 211–224 (2004)
36. N Dalal, B Triggs. Histograms of oriented gradients for human detection. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **1**, 886–893 (2005)
37. YD Sato, C Wolff, P Wolfrum, C von der Malsburg. Dynamic link matching between feature columns for different scale and orientation, in *Neural Information Processing*, ed. by M Ishikawa, K Doya, H Miyamoto, T Yamakawa. 14th International Conference, ICONIP 2007, Kitakyushu, Japan, November 13–16, 2007, Revised Selected Papers, Part I. Lecture Notes in Computer Science, vol. 4984 (Springer Berlin, 2007), pp. 385–394
38. YD Sato, J Jitsev, C von der Malsburg. A visual object recognition system invariant to scale and rotation. *Neural Netw. World.* **19**(5), 529–544 (2009)
39. N Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* **9**(1), 62–66 (1979)
40. B Sirmacek, C Ünsalan. Building detection using Gabor features in very high resolution satellite images, in *4th International Conference on Recent Advances in Space Technologies (RAST)* (IEEE Piscataway, 2009)
41. T Nakano, T Morie. A digital LSI architecture of elastic graph matching and its FPGA implementation, in *Proceedings of the International Joint Conference on Neural Networks (IJCNN05)* (IEEE Piscataway, 2005), pp. 689–694

doi:10.1186/1687-6180-2013-175

**Cite this article as:** Sato and Kuriya: Multi-scale elastic graph matching for face detection. *EURASIP Journal on Advances in Signal Processing* 2013 **2013**:175.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)