World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:6, No:3, 2012

Efficient and Effective Gabor Feature Representation for Face Detection

Yasuomi D. Sato, Yasutaka Kuriya

Abstract**—**We here propose improved version of elastic graph matching (EGM) as a face detector, called the multi-scale EGM (MS-EGM). In this improvement, Gabor wavelet-based pyramid reduces computational complexity for the feature representation often used in the conventional EGM, but preserving a critical amount of information about an image. The MS-EGM gives us higher detection performance than Viola-Jones object detection algorithm of the AdaBoost Haar-like feature cascade. We also show rapid detection speeds of the MS-EGM, comparable to the Viola-Jones method. We find fruitful benefits in the MS-EGM, in terms of topological feature representation for a face.

Keywords**—**Face Detection, Gabor Wavelet Based Pyramid, Elastic Graph Matching, Topological Preservation, Redundancy of Computational Complexity.

I. Introduction

FACE detection is one of important processes in computer vision technology. It is comparable to face recognition. Many different types of face detection algorithms have so far been suggested and developed actively towards practical applications [1]. Such face detection algorithms have been comparatively tested [2]. One of them, which is remarkably exploited in recent computer visions, is a cascade of boosted classifiers with Haar-like (HL) features, proposed by Viola and Jones [3, 4]. In the Viola-Jones algorithm, a set of weak classifiers are used. The large number of weak classifiers are singled out and then organized into cascade. Viola and Jones succeeded to realize rapid face detection. Face recognition itself is a core task in system developments for face recognition. Many algorithms for face recognition were developed. One of them is well-known widely as face recognition literature, which is called the Elastic Graph Matching (EGM) [5] or the Elastic Bunch Graph Matching (EBGM) [6]. Their essential concept for effective facial identification is so-called Dynamic Link Architecture (DLA) [7] that Gabor feature detectors establish dynamic projections of an undistorted graph to the elastic one.

Y. D. Sato is with Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4, Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan (corresponding author to provide phone: +81 93-695-6234; fax: +81 93-695-6234; e-mail: sato-y@ brain.kyutech.ac.jp). Frankfurt Institute for Advanced Studies (FIAS), Johann Wolfgang Goethe University, Ruth-Moufang-Str. 1, D60438, Frankfurt am Main, Germany. Brain Science Institute, RIKEN, 2-1, Hirosawa, Wako, Saitama, 351-0198, Japan.

Y. Kuriya was with Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4, Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan (e-mail: kuriya@brain.kyutech.ac.jp).

In the algorithm, there seems to still exist one serve problem, that is, no face detection functionality. Wolfrum et al. suggested a recurrent network model for processing of both detection and identification of a face on an input image [8]. The network model must be very interesting, but its technical improvements are furthermore requested toward practical applications.

For this purpose, the current urgent task is to progressively develop EGM-based face detection algorithm, which possibly has a comparable capability for face detection to the AdaBoost of the HL cascade type. In this work, we suggest a face detection algorithm developed with skillful combination of EGM and image pyramids. In what follows, we call it as the multi-scale EGM (MS-EGM).

In the MS-EGM, efficient and effective feature representation for a face is employed with Gabor wavelet based pyramid [9]. This effectively reduces computational complexity of the feature representation, preserving critical amount of representational information about the face, in comparison with conventional Gabor feature representations. Such efficient and effective representation redundancy may bring about a more simplified system for face detection.

II. Multi Scale Elastic Graph Matching

A. System Configuration

We shall begin by explaining a whole configuration of a face detector with multi scale elastic graph matching (MS-EGM). The system consists basically of preprocess for down-sample and then Gabor wavelet transform (GWT) and main process for face detection with EGM.

B. Gabor Pyramid in Model

We employ an image of the average face, whose size is A_0=60 $\times$60 pixels. Since we suppose feature representations being composed of the different scales. The image is down-sampled with a_0^s (s=0, …, 5), called the M_s. Here a_0=0.85 and s=0 implies the original size of the image M. A square graph of ($n \times n$ - 4) nodes without any vertexes is set on each resolution image M_s. Here n (=5) is the number of full nodes on a row and column of the square graph. Each node on the image M_s is convoluted with a family of Gabor functions $\Psi_r(z)$. r (=0, …, 7) is an orientation parameter. The Gabor feature usually consists of 8 different orientation components. Each orientation component is given by a convoluted value ($\hat{J}_r^{M_s}$):

$$\hat{J}_r^{M_s}(p_s) = \int F^{M_s}(p_s - \hat{p}_s)\Psi_r(p_s - \hat{p}_s)d^2\hat{p}_s\,, \tag{1}$$

$$\psi_r(p_s) = \frac{k_r^2}{\sigma^2} \exp\left(-\frac{k_r^2 p_s^2}{2\sigma^2}\right)\left[\exp(ik_r p_s) - \exp\left(-\frac{\sigma^2}{2}\right)\right]. \tag{2}$$

$\sigma=2\pi$. The wave number vector k_r can be expressed as

$$\vec{k}_r = \begin{pmatrix} k_{rx} \\ k_{ry} \end{pmatrix} = \begin{pmatrix} k\cos\varphi_r \\ k\sin\varphi_r \end{pmatrix}, \quad k = \frac{\pi}{2}, \quad \varphi = \frac{\pi}{8} r \cdot \tag{3}$$

One orientation component $J_r^{M_s}$ in the Gabor feature takes an absolute value of $\hat{J}_r^{M_s}$:

$$J_r^{M_s}(p_s) = \left|\hat{J}_r^{M_s}(p_s)\right| \cdot \tag{4}$$

C. Face Detection

In analogy, the input (I) image is convoluted with a family of Gabor function. We then proceed main face detection processing. The face detection processing consists of two sub-processes of candidate position finding and the most likely position detection.

1. Scan Matching

In the first sub-process, in order to pick up candidate positions, an entire or fragmentary similarity map for each scale M_s is calculated with scan matching of the undistorted graph G_s onto the image I:

$$e_s^S(x_I) = \frac{1}{|G_s|} \sum_{p_s \in G_s} \frac{\sum_r J_r^I(x_I) \cdot J_r^{M_s}(p_s)}{\sqrt{\sum_r \left(J_r^I(x_I)\right)^2 \sum_r \left(J_r^{M_s}(p_s)\right)^2}}, \tag{5}$$

where $e_s^S(x_I)$ is the similarity of the Gabor feature at x_I on the image I to the one for each node p_s. It is noticed that G_s sometimes represents a set of nodes of the graph. In the scan matching process, the left-upper of the image M_s on which a square graph is being set up, is firstly adjusted to the left-upper of the I. The M_s is repeatedly scanned to the right for each row on the I. The scan ends when the right-lower of the M_s arrives at the right-lower of the I (not shown here). As a result, we obtain a similarity map at each image M_s. We exemplify similarity map results when one facial image I is scanned with a square graph at the level M_5.

Next, we pick up some candidate positions that can be expected as the one of a face. Here let the candidate position $\mathbf{x}_I^c$ be defined as the center of the image M_s or of its square graph. It corresponds to the local maximum on the similarity map. The local maximum is satisfied with a gradient condition when all differences of the $\mathbf{x}_I^c$ to its nearest neighbors $\mathbf{x}_I^n$ take positive:

$$e_s^S(x_I^c) - e_s^S(x_I^n) > 0\,. \tag{6}$$

The candidate pixel is depicted with a black square on the similarity map.

2. *Elastic Graph Matching (EGM)*

The EGM for each candidate position $\mathbf{x}_I^c$ is computed to obtain the maximum value of the cost function E_s^c, which is given by

$$E_s^c(x_I) = e_s^S(x_I) - \lambda_d e_s^d(x_I)\,, \tag{7}$$

$$e_s^d(x_I) = \frac{1}{|G_s|} \sum_{p_s \in G_s} \frac{\left|\sum_{p'_s \in G'_s} D_{p_s\,p'_s}^{M_s} - \sum_{p'_s \in G'_s} D_{x_I\,p'_s}^{I}\right|}{\left|\sum_{p'_s \in G'_s} D_{p_s\,p'_s}^{M_s}\right|} + 1 - \frac{1}{|G_s|} \sum_{p_s \in G_s} \frac{\left|\sum_{p'_s \in G'_s} A_{p_s\,p'_s}^{M_s} \cdot \sum_{p'_s \in G'_s} A_{x_I\,p'_s}^{I}\right|}{\left|\sum_{p'_s \in G'_s} A_{p_s\,p'_s}^{M_s}\right| \cdot \left|\sum_{p'_s \in G'_s} A_{x_I\,p'_s}^{I}\right|}, \tag{8}$$

where $e_s^d(x_I)$ represents the elasticity of the graph on the image I. λ_d is a constant parameter for the graph elasticity. λ_d=0.05, except for obtaining a similarity map $E_s(x_I)$ when λ_d=0. G'_s is a set of nearest neighbor nodes p'_s for p_s. $D_{p_s\,p'_s}^{M_s}$ and $D_{x_I\,p'_s}^{I}$ are the Euclidean distance between nodes p_s (or x_I) and p'_s on the graph of the image M_s (or I). $A_{p_s\,p'_s}^{M_s}$ and $A_{x_I\,p'_s}^{I}$ take one vector form consisting of 4 elements. Each element is an angular between two nearest neighbors on each quadrant, centered at p_s.

Each node on the image I, which corresponds to the node p_s^c of the square graph M_s, surveys an optimal pixel p_s^I taking a maximum of the cost function E_s^c within a search region R:

$$p_s^I = \max_{x_I \in R} \left\{E_s^c(x_I)\right\} \tag{9}$$

where R is a set of pixels that can be picked up in a square with the size of ($2q$+1)×($2q$+1), centered at the pixel corresponding to p_s^c. q=4.

Finally, the optimal pixel x_m^I is singled out with the maximum operation of all candidates, which must be the most invariant to feature representation for an M face:

$$x_m^I = \max_{c \in C} \left\{E_s^c(p_s^I)\right\}. \tag{10}$$

Here C is a set of candidates.

III. Ability Test

We test a face detection ability of multi-scale elastic graph matching (MS-EGM), using a database of the BioID (Fig.1) [10], compared to an AdaBoost type of the HL cascade face detector developed by Viola and Jones [3, 4], included by default in OpenCV 2.0: *haarcascade_ontalface_alt* trained by Lienhart et al. [11]. An image of the average face created with many German facial photos will be employed as the Model (M), on which sets a square graph. Both the face detectors are simulated on the same computers. We assume correct detection be defined when eyes are located in three-fifths from the top of obtained square area whereas the incorrect means that the eyes' position is out of the upper area. We show the face detection performance result in Table I. In Table I, our MS-EGM has shown the higher correct detection rate of 97.6% than the AdaBoost HL cascade. Their computational cost per one facial image is almost same. This face detection result may suggest that our MS-EGM has the same or rather better ability for face detection, compared to the AdaBoosted HL cascade. However we may have to pay an attention to the face detection ability tests. In this work, we used only one database. It will be necessary to test the face detection abilities with the other databases such as FERET, Caltech Faces and AR Faces. We should then evaluate the ability of our MS-EGM for face detection with the overall estimation.

IV. DISCUSSION

In this work, we have proposed a face detector that should be worthy of comparison to detection capability of the AdaBoost HL cascade type, from results of ability tests. However, it may still request the higher performance for the sake of its effective works for real. It would thus be necessary for us to inspect causes of false detection in the experiments, or drawbacks in our system.Most of the false detections were detecting the background on the input image. The reason why detecting the background is the Gabor feature representation with a square grid graph for a model face. Most of grids are located on fiducial points on a face like eyebrows, eyes, lips and facial edge. Such square graph-like feature representations are significantly sensitive to linear edges observed in the backgrounds. Indeed, they take higher similarities to representations for linear edges, compared to the facials. We have found that the square-like feature representation elicits misrecognition of the round-necked (or lips) to lips (or eyes) as well. In order to solve to such a misrecognition problem, the right way is to use a face graph instead of a square graph. It can be distinguished against the background to be less misrecognition.The other reason of the misdetection is that the size of an 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 trade-off between computational costs and the number of the model face size to be prepared, because 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.Next, we consider a reason why our MS-EGM could be capable of as rapid image processing as the AdaBoost HL cascade. The largest cause is full redundancy of computational complexity in Gabor feature representation, but preserving a critical amount of information about a face. In conventional EGM, facial images were expressed with full Gabor feature representations encompassing many different orientation and many different spatial frequencies of the Gabor wavelet kernel [12]. This induces consuming time increments. However in our MS-EGM, the Gabor wavelet Pyramid, employed for finding multi-scale feature correspondence [9], enables us to remove unnecessary Gabor filter, in particular, with the low spatial frequency of the kernel, which increases the consuming time.

The same was in the scale invariant feature transform (SIFT) method [13, 14]. In the SIFT, a Gaussian pyramid is used for ruling out the undesired filtering with the large kernel. However the Gaussian pyramid apparently looses much information about a face. This means to difficultly preserve such information in the feature representation. Thus, the SIFT has to be incorporated with additional feature representation such as a histogram of oriented-gradient (HoG). Such duplicate feature representation have just increased the computational complexity conversely, even though the SIFT with the HoG feature might be logically reasonable.

Fig. 1 Facial images of a same or different person with the background in a Bio ID dataset of 1521 images

TABLE I
FACE DETECTION PERFORMANCE

Database	MS-EGM Rate (%), Cost (ms)	AdaBoosted HL Cascade Rate (%), Cost (ms)
BioID	97.6, 37	96.2, 36

Face detection ability tests of the MS-EGM and the AdaBoosted HL cascade, by using the database of the BioID with 1521 facial images.

Therefore, we can address that even in comparison with the SIFT, our MS-EGM has a lot of advantages.

Finally, we look for more advantages of the MS-EGM, in terms of topological feature representation. The greatest benefit in the MS-EGM is the topological preservation, based on correspondence finding between the input and model images. This means that the MS-EGM has the strong robustness against distortion invariance, including occlusions. It expects that the MS-EGM can effortlessly detect a face on its partially occluded image. In contrast, the AdaBoost HL cascade can difficultly detect a face on the occluded image, because it naturally has no topological preservation in the feature presentation. We must create another cascade of boosted classifiers, strongly robust to faces occluded with sunglasses or scarf.

In any case, we have proposed a novel EGM for achieving effective face detection, called the MS-EGM. The greatest improvement in the conventional EGM is effective and efficient feature representation with Gabor wavelet based pyramid, in comparison with the SIFT with the HoG features. We also show that the MS-EGM has a higher face detection rate with comparably rapid image processing, comparing to the Adaboost HL cascade. In discussion, we have expected the MS-EGM is strong robust to occluded face images.

ACKNOWLEDGMENT

This work was supported in part by the Grant-in-Aid for Young Scientist (B) No. 22700237. Authors thank von der Malsburg for fruitful and active discussion. von der Malsburg was a coordinator of the “Bernstein Focus: Neurotechnology through research grant 01GQ0840.”

REFERENCES

[1] E. Hjelmas, B. K. Low, “Face Detection: A Survey.” Computer Vision and Image Understanding, 83(3), pp. 236–274, 2001.

[2] N. Degtyarev, O. Seredin, “Comparative Testing of Face Detection Algorithms.” In A. E lmoataz et al. (Eds.): ICISP 2010, LNCS 6134, pp. 200–209 , (Springer-Verlag, Heidelberg, 2010).

[3] P. Viola, M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple." Proc. IEEE CVPR 2001, 2001.
[4] P. Viola, M. J. Jones, "Robust Real-Time Face Detection." International Journal of Computer Vision, 57(2), pp.137–154, 2004.
[5] M. Lades, J. C. Vorbrueggen, J. Buhmann, et al., "Distortion invariant object recognition in the dynamic link architecture." IEEE Transactions on Computers, vol.42, no. 3, pp. 300–311, 1993.
[6] L. Wiskott, J.-M. Fellous, N. Kruger, C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching." IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7), pp. 775–779, 1997.
[7] C. von der Malsburg, "The correlation theory of brain function." Internal Report Max-Planck-Institute fur Biophysikalische Chemie, Postfach 2841, D-3400 Gottingen, FRG, 1981.
[8] P. Wolfrum, C. Wolff, J. Lucke, C. von der Malsburg, "A Recurrent Dynamic Model for Correspondence-Based Face Recognition." Journal of Vision, vol.8, pp.1–18, 2008.
[9] Y. D. Sato, J. Jitsev, J. Bornschein, D. Pamplona, C. Keck, C. von der Malsburg, "A Gabor Wavelet Pyramid-Based Object Detection Algorithm." In Proc. ISNN (2) pp.232–240, 2011.
[10] O. Jesorsky, K. Kirchberg, R. Frischholz, "Robust Face Detection Using the Hausdorff Distance." In J. Bigun and F. Smeraldi, editors, Audio and Video based Person Authentication - AVBPA 2001, Springer, pp.90–95, 2001.
[11] R. Lienhart, J. Maydt, "An extended set of Haar-like features for rapid object detection." Proc. Intern. Conf. on Image Processing 2002, pp.900–903, 2002.
[12] A. Eleyan, H. Ozkaramanli, H. Demirel, "Complex wavelet transform-Based face recognition." EURASIPJournal on Advances in Signal Processing, vol. 2008, Article ID 185281, 13 pages, 2008.
[13] D. G. Lowe, "Object Recognition from Local Scale-Invariant Features." Proceedings of the International Conference on Computer Vision 1999, 2, pp.150–1157, 1999.
[14] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints." International Journal of Computer Vision, 60(2) pp. 91–110, 2004.

Yasuomi D. Sato (Fujieda, Shizuoka, Japan and 25th June 1974) received the Doctor of Science and M.S. degrees from Department of Physics (Condensed Matter Physics), Tokyo Institute of Technology, Tokyo, Japan, respectively in 2005 and 2001. He received the B.S. degree from Saitama University, Saitama, Japan in 1998. He served as Junior Fellow (corresponding to Junior Professor) at Frankfurt Institute for Advanced Studies (FIAS), Frankfurt am Main, Germany since 2005. He was engaging in European Commission-funded project, FP6-2005-015803. He was Assistant Professor at Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology since 2009.

Yasutaka Kuriya (Onga, Fukuoka, Japan and 2nd February 1981) received Master of Engineering degree from Department of Brain Science and Engineering, Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Fukuoka, Japan in 2009. He received the B.E. degree from Kitakyushu National College of Technology, Fukuoka, Japan in 2007.