BAYESIAN FACE DETECTION IN AN IMAGE SEQUENCE USING FACE PROBABILITY GRADIENT ASCENT

Jae Hee Park, Hae Chul Choi**, and Seong Dae Kim**

*Div. of EE, Dep. of EECS, Korea Advanced Institute of Science and Technology, Daejeon, Korea
** Digital Broadcasting Research Division, Electronics and Telecommunications Research Institute
E-mail: *{yscclo, sdkim}@sdvision.kaist.ac.kr, **choihc@etri.re.kr

ABSTRACT

Face detection in an image sequence is a challenging problem for many applications. In this paper, a novel face detection method is proposed. In order to detect faces in a sequence, based on Bayesian decision theory, we construct a unified framework of most face-like region selection, face/non-face classification, and detection result correction. And we propose Face Probability Gradient Ascent method to estimate the optimal position, scale, and rotation parameters of each face. In the experimental results, it is shown that the proposed method is more accurate and efficient than other conventional detection methods.

1. INTRODUCTION

Face detection is the process to determine whether one or more human faces exist in one or more images and to find the faces' locations. Face detection is a basic and important problem for many applications such as visual surveillance, HCI (Human Computer Interface) and so on.

It is one of the most challenging problems to find faces which have variations of scale and rotation [1][3][4]. In order to solve the problem, scale or rotation invariant features can be used [4]. However, considering information loss, to compensate the variations and extract feature from the compensated image is better, if possible [3]. In that case, the detection performance largely depends on the criterion for deciding scale or rotation parameters.

Face detection in an image sequence is a widely researched subject nowadays [1][3]. A merit of using a sequence instead of one image is possibility to correct the wrong results of each frame-wise detection procedure [1].

The proposed method in this paper is a face detection method based on Bayesian decision theory, which uses the whole sequence concurrently.

There are two main contributions of the proposed method.

The first contribution is a Bayesian formulation for an image sequence. We propose a Bayesian formulation and construct a unified framework of most face-like region selection, face/non-face classification, and detection result correction for sequence-wise face detection.

The second contribution of this paper is a warping parameter estimation method called FPGA (Face Probability Gradient Ascent). It allows fast and accurate warping parameter estimation for locating faces which have variations of scale and rotation.

The detailed description of the proposed method is as follows.

2. THE PROPOSED DETECTION METHOD

2.1. Basic Assumption

The n-th input image of the sequence is denoted as a one-dimensional vector as eq. (1).

$$\mathbf{i}(n) = \left(i_1(n), i_2(n), \cdots i_{XY}(n)\right)^T \quad (1)$$

Assume that there is a human face in the predefined region of the image and that the scale and rotation of the face agree with those of training faces. In that case, there exists a fixed matrix $\mathbf{F} = \{f_{ij} \mid f_{ij} = 0 \text{ or } 1\}$ such that the partial image $\mathbf{Fi}(n)$ agrees with the face region.

Generally, position, scale and rotation of the contained face disagree with those of training faces. Then, there exist appropriate parameters $\mathbf{c}(n) = [c_x(n), c_y(n)]^T$, $s(n)$, $\theta(n)$ which correspond to translation, scale, and rotation such that the partial image $\mathbf{F}\tilde{\mathbf{i}}(n)$ agrees with the face region where $\tilde{\mathbf{i}}(t)$ is the warped image by the parameters.

The proposed method in this paper is based on $\Pr\left(\omega_f \mid \mathbf{F}\tilde{\mathbf{i}}(n)\right)$ which is the probability that the partial image $\mathbf{F}\tilde{\mathbf{i}}(n)$ corresponds to a human face.

2.2. Bayesian Formulation for Detection in a Sequence

In this paper, an image sequence in the interval $0 \le n \le N$ is denoted as $\mathbf{I} = \{\mathbf{i}(n) \mid n = 0,1,\cdots,N\}$. Similarly, $\boldsymbol{\Phi} = \{\boldsymbol{\varphi}(n) \mid n = 0,1,\cdots,N\}$ denotes a set of parameter vectors $\boldsymbol{\varphi}(n) = \left[s(n), \theta(n), c_x(n), c_y(n) \right]^{\mathrm{T}}$.

If there exist probability models of $p(\boldsymbol{\Phi})$ and $\Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi})$, the probability that $\mathbf{I}$ contains a face can be calculated as an expectation of $\Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi})$.

$$\Pr(\omega_f \mid \mathbf{I}) = \sum_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}) \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) \quad (2)$$

where $p(\boldsymbol{\Phi})$ is probability density function of $\boldsymbol{\Phi}$. In this paper, eq. (3) replaces eq. (2) for the sake of computational convenience.

$$\Pr(\omega_f \mid \mathbf{I}) = \max_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}) \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) \quad (3)$$

This replacement is based on one or both of following assumptions.

i. $\max_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}) \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) \propto \sum_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}) \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi})$

ii. If $\mathbf{F}\tilde{\mathbf{i}}(n)$ is not well aligned with human faces, $\Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) \simeq 0$

2.3. Face Probability Evaluation

For a given image and a parameter set, the probability of face existence at the predefined region can be evaluated as follows.

$$\begin{aligned} \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) &\propto p(\mathbf{I}, \boldsymbol{\Phi} \mid \omega_f) = p(\mathbf{F}\tilde{\mathbf{i}}(0), \mathbf{F}\tilde{\mathbf{i}}(1), \cdots, \mathbf{F}\tilde{\mathbf{i}}(N) \mid \omega_f) \\ &= p(\mathbf{F}\tilde{\mathbf{i}}(0) \mid \omega_f) p(\mathbf{F}\tilde{\mathbf{i}}(1) \mid \omega_f) \cdots p(\mathbf{F}\tilde{\mathbf{i}}(N) \mid \omega_f) \end{aligned} \quad (4)$$

under an assumption $p(\mathbf{F}\tilde{\mathbf{i}}(m) \mid \omega_f, \mathbf{F}\tilde{\mathbf{i}}(n)) = p(\mathbf{F}\tilde{\mathbf{i}}(m) \mid \omega_f)$ for $m \neq n$. In this paper, the conditional probability density function $p(\mathbf{F}\tilde{\mathbf{i}}(n) \mid \omega_f)$ is obtained using Liu's method [2] with some simplifications. Based on the method, we extract feature vectors by a simple linear calculation and approximate the conditional *pdf* as a multivariate Gaussian distribution. The feature vector $\mathbf{z}$ is extracted from $\mathbf{F}\tilde{\mathbf{i}}(t)$ by the method as eq. (5).

$$\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}) = \boldsymbol{\Phi}_f^T (\mathbf{A}\mathbf{F}\tilde{\mathbf{i}} - \mathbf{m}_f) = \mathbf{M}\tilde{\mathbf{i}} + \mathbf{v} \quad (5)$$

where $\mathbf{A}$ is a matrix for feature extraction [2], $\mathbf{m}_f$ is the mean of feature vectors $\mathbf{A}\mathbf{i}_{tr}$ from training face regions $\mathbf{i}_{tr}$, and each column of $\boldsymbol{\Phi}_f$ is an eigenvector of $\mathbf{A}\mathbf{i}_{tr} - \mathbf{m}_f$.

By eq. (5), the feature vector $\mathbf{z}$ is extracted and $p(\mathbf{F}\tilde{\mathbf{i}} \mid \omega_f) = p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}) \mid \omega_f)$ is calculated.

2.4. Probability Model for Parameter Set

Probability density function $p(\boldsymbol{\Phi})$ is can be considered as eq. (6) by the chain rule and assuming $\boldsymbol{\varphi}(n)$ as a 1st order Markov random process.

We construct a probability model as eq. (7) assuming that parameters from neighboring frames are similar.

$$p(\boldsymbol{\Phi}) = p(\boldsymbol{\varphi}(N) \mid \boldsymbol{\varphi}(N-1)) \cdots p(\boldsymbol{\varphi}(1) \mid \boldsymbol{\varphi}(0)) \cdot \Pr(\boldsymbol{\varphi}(0)) \quad (6)$$

$$p(\boldsymbol{\varphi} \mid \boldsymbol{\varphi}(n-1)) = \frac{1}{\sigma_{\boldsymbol{\varphi}} (2\pi)^2} \exp\left\{ -\frac{1}{2\sigma_{\boldsymbol{\varphi}}^2} \left\| \boldsymbol{\varphi} - \boldsymbol{\varphi}(n-1) \right\|^2 \right\} \quad (7)$$

2.5. Optimal Parameter Estimation

Eq. (3) can be considered as eq. (8) by substituting eq. (4) and eq. (6).

$$\begin{aligned} \Pr(\omega_f \mid \mathbf{I}) &= \max_{\boldsymbol{\Phi}} p(\boldsymbol{\Phi}) \Pr(\omega_f \mid \mathbf{I}, \boldsymbol{\Phi}) \\ &\propto \max_{\boldsymbol{\Phi}} \prod_{n=1}^{N} \left(p(\boldsymbol{\varphi}(n) \mid \boldsymbol{\varphi}(n-1)) \, p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}(n)) \mid \omega_f) \right) \\ &= \prod_{n=1}^{N} \max_{\boldsymbol{\varphi}(n)} \left(p(\boldsymbol{\varphi}(n) \mid \boldsymbol{\varphi}(n-1)) \, p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}(n)) \mid \omega_f) \right) \end{aligned} \quad (8)$$

From this equation, for the given $\boldsymbol{\varphi}(n-1)$, the optimal parameter vector $\hat{\boldsymbol{\varphi}}(n)$ at time n is defined as eq. (9).

$$\hat{\boldsymbol{\varphi}}(n) = \arg\max_{\boldsymbol{\varphi}(n)} p(\boldsymbol{\varphi}(n) \mid \boldsymbol{\varphi}(n-1)) \, p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}(n)) \mid \omega_f) \quad (9)$$

In this paper, in order to find the optimal parameters which maximizes $p(\boldsymbol{\varphi}(n) \mid \boldsymbol{\varphi}(n-1)) \, p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}(n)) \mid \omega_f) = p(\boldsymbol{\varphi}) \, p(\mathbf{z} \mid \omega_f)$, Face Probability Gradient Ascent (FPGA) method is proposed. The proposed FPGA method is a gradient-based iterative method of which object function is a probability density function.

At the k-th iteration, a new parameter vector is obtained as eq. (10).

$$\boldsymbol{\varphi}_k = \boldsymbol{\varphi}_{k-1} + \boldsymbol{\Delta\varphi} = \boldsymbol{\varphi}_{k-1} + \alpha \nabla_{\boldsymbol{\varphi}} \left(p(\boldsymbol{\varphi}) \, p(\mathbf{z} \mid \omega_f) \right) \Big|_{\boldsymbol{\varphi} = \boldsymbol{\varphi}_{k-1}} \quad (10)$$

where α is called step-size.

The gradient of $p(\boldsymbol{\varphi}) \, p(\mathbf{z} \mid \mathrm{Face})$ for $\boldsymbol{\varphi}$ at $\boldsymbol{\varphi} = \boldsymbol{\varphi}_{k-1}$, $\nabla_{\boldsymbol{\varphi}} \left(p(\boldsymbol{\varphi}) \, p(\mathbf{z} \mid \omega_f) \right) \Big|_{\boldsymbol{\varphi} = \boldsymbol{\varphi}_{k-1}}$, can be calculated as follows.

$$\nabla_{\boldsymbol{\varphi}}\left(p(\boldsymbol{\varphi})\,p(\mathbf{z}\,|\,\omega_f)\right)\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_{k-1}}$$
$$= p(\mathbf{z}\,|\,\omega_f)\nabla_{\boldsymbol{\varphi}} p(\boldsymbol{\varphi})\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_{k-1}} + p(\boldsymbol{\varphi}_{k-1})\nabla_{\boldsymbol{\varphi}} p(\mathbf{z}\,|\,\omega_f)\Big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_{k-1}} \tag{11}$$

$\nabla_{\boldsymbol{\varphi}} p(\mathbf{z}\,|\,\omega_f)$ of eq. (11) can be considered as multiplication of two parts as eq. (12).

$$\nabla_{\boldsymbol{\varphi}} p(\mathbf{z}\,|\,\omega_f) = \frac{\partial \mathbf{z}}{\partial \boldsymbol{\varphi}} \frac{\partial p(\mathbf{z}\,|\,\omega_f)}{\partial \mathbf{z}} \tag{12}$$

Considering eq. (13), $\frac{\partial p(\mathbf{z}\,|\,\omega_f)}{\partial \mathbf{z}}$ can be calculated as eq. (14).

$$p(\mathbf{z}\,|\,\omega_f) = \frac{1}{(2\pi)^{M/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{z}-\mathbf{m}_{\mathbf{x}})^T \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\mathbf{m}_{\mathbf{x}})\right\} \tag{13}$$

$$\frac{\partial p(\mathbf{z}\,|\,\omega_f)}{\partial \mathbf{z}} = -p(\mathbf{z}\,|\,\omega_f) \times \boldsymbol{\Sigma}^{-1}(\mathbf{z}-\mathbf{m}_{\mathbf{x}}) \tag{14}$$

where M , $\mathbf{m}_{\mathbf{x}}$, $\boldsymbol{\Sigma}$ are feature dimension, the mean feature vector of training faces, and the covariance matrix of features respectively.

Considering eq. (5), $\frac{\partial \mathbf{z}}{\partial \boldsymbol{\varphi}}$ can be calculated as eq. (15).

$$\frac{\partial \mathbf{z}}{\partial \boldsymbol{\varphi}} = \frac{\partial (\mathbf{M}\tilde{\mathbf{i}}_{k-1} + \mathbf{v})}{\partial \boldsymbol{\varphi}} = \begin{pmatrix} \frac{\partial \tilde{i}_{k-1,1}}{\partial s} & \frac{\partial \tilde{i}_{k-1,1}}{\partial \theta} & \frac{\partial \tilde{i}_{k-1,1}}{\partial c_x} & \frac{\partial \tilde{i}_{k-1,1}}{\partial c_y} \\ \vdots & & & \vdots \\ \frac{\partial \tilde{i}_{k-1,N}}{\partial s} & \frac{\partial \tilde{i}_{k-1,N}}{\partial \theta} & \cdots & \frac{\partial \tilde{i}_{k-1,N}}{\partial c_y} \end{pmatrix}^T \mathbf{M}^T \tag{15}$$

where $\tilde{i}_{k-1,n}$ is the n -th element of the warped image vector at $(k-1)$ -th iteration.

In this paper, we assume that $\theta \approx 0, s \approx 1$ and $\tilde{i}_{k-1,n}$ corresponds to the position $\mathbf{x} = (x_0, y_0)^T$. Then, each element $\partial \tilde{i}_{k-1,n} / \partial s$ can be calculated as eq. (16) by eq. (17) and eq. (18).

$$\frac{\partial \tilde{i}_{k-1,n}}{\partial s} = \frac{\partial \tilde{i}_{k-1,n}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial s} = x_0 \frac{\partial \tilde{i}_{k-1,n}}{\partial x}\bigg|_{(x_0,y_0)} + y_0 \frac{\partial \tilde{i}_{k-1,n}}{\partial y}\bigg|_{(x_0,y_0)} \tag{16}$$

$$\frac{\partial \tilde{i}_{k-1,n}}{\partial \mathbf{x}} = \nabla \tilde{i}_{k-1,n} = \left(\frac{\partial \tilde{i}_{k-1,n}}{\partial x} \quad \frac{\partial \tilde{i}_{k-1,n}}{\partial y}\right)\bigg|_{(x_0,y_0)} \tag{17}$$

$$\frac{\partial \mathbf{x}}{\partial s} = \begin{pmatrix} \partial x / \partial s \\ \partial y / \partial s \end{pmatrix} = \begin{pmatrix} x_0 \cos\theta - y_0 \sin\theta \\ x_0 \sin\theta + y_0 \cos\theta \end{pmatrix} \simeq \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \tag{18}$$

Similarly, we can get eq. (19) ~ eq. (21).

$$\frac{\partial \tilde{i}_{k-1,n}}{\partial \theta} \simeq -y_0 \frac{\partial \tilde{i}_{k-1,n}}{\partial x}\bigg|_{(x_0,y_0)} + x_0 \frac{\partial \tilde{i}_{k-1,n}}{\partial y}\bigg|_{(x_0,y_0)} \tag{19}$$

$$\partial \tilde{i}_{k-1,n} / \partial c_x = \partial \tilde{i}_{k-1,n} / \partial x\big|_{(x_0,y_0)} \tag{20}$$

$$\partial \tilde{i}_{k-1,n} / \partial c_y = \partial \tilde{i}_{k-1,n} / \partial y\big|_{(x_0,y_0)} \tag{21}$$

$\nabla_{\boldsymbol{\varphi}} p(\boldsymbol{\varphi})\big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_{k-1}}$ of eq. (11) can be calculated as eq. (22).

$$\nabla_{\boldsymbol{\varphi}} p(\boldsymbol{\varphi})\big|_{\boldsymbol{\varphi}=\boldsymbol{\varphi}_{k-1}} = -1/\sigma_{\boldsymbol{\varphi}}^{2} \times p(\boldsymbol{\varphi}_{k-1}) \times (\boldsymbol{\varphi}_{k-1} - \boldsymbol{\varphi}(n-1)) \tag{22}$$

From this procedure, $\boldsymbol{\varphi}_k = \boldsymbol{\varphi}_{k-1} + \boldsymbol{\Delta\varphi}$ is obtained.

In this paper, we repeat the calculation of eq. (11) and get $\hat{\boldsymbol{\varphi}}(t)$.

2.6. Multiple Face Detection Strategy

Face detection algorithm using the proposed method is as follows.

step 1. For the initial input image $\mathbf{i}(0)$, detect faces using Liu's method [2].

step 2. Calculate the initial parameters $\boldsymbol{\varphi}_0^k(0)$ so that the k -th face region corresponds to $\mathbf{F}\tilde{\mathbf{i}}(0)$.

step 3. For $\boldsymbol{\varphi}_0^k(0)$ which corresponds to each face region, get the optimal parameters $\hat{\boldsymbol{\varphi}}^k(0)$ using FPGA method proposed in section 2.5. Then, compose the face list with calculated $\hat{\boldsymbol{\varphi}}^k(0)$ and evaluate $p(\mathbf{F}\tilde{\mathbf{i}}(0)\,|\,\omega_f)$.

step 4. For the input image $i(n)$ at time n , set $\boldsymbol{\varphi}_0^k(n) = \hat{\boldsymbol{\varphi}}^k(n-1)$, calculate $\hat{\boldsymbol{\varphi}}^k(n)$, update the face list, and evaluate $\Pr(\mathbf{F}\tilde{\mathbf{i}}(n)\,|\,\omega_f)$.

step 5. For each frame, calculate $\prod_{m=1}^{n} \max_{\boldsymbol{\varphi}(m)} \left(p(\boldsymbol{\varphi}(m)\,|\,\boldsymbol{\varphi}(m-1))\,p(\mathbf{z}(\mathbf{F}\tilde{\mathbf{i}}(m))\,|\,\omega_f)\right)$. And if the value is below th^n , remove the face from the face list.

step 6. At every 10 frames, repeat step 1, and modify the face list.

From this procedure, we can detect multiple faces even if some faces appear or disappear in the middle of the image sequence. Furthermore, we can correct the initial detection failures or false alarms.

3. EXPERIMATAL RESULTS

3.1. Locating Face using FPGA

Using FPGA method proposed in this paper, for a given image, we got the most face-like region near the given initial face.

In the first column of Fig. 1(a), the initial face regions given manually are shown and in other columns, there are the results at 5-th and 15-th iteration respectively. Fig. 1(b) shows face likelihood $p(\mathbf{z} \mid \omega_f)$ at each iteration step.

In experiments, we observed that face likelihood largely depends on the accuracy of face region selection and we got very accurate face regions using the FPGA method.

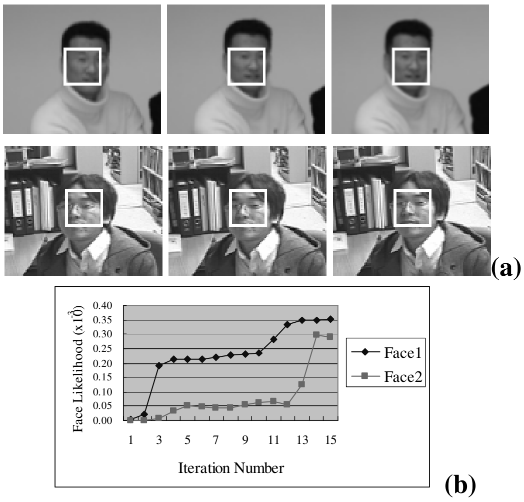

Fig. 1 Performance of FPGA Method in an Image

3.2. Face Detection in an Image Sequence

We implemented a face detection system for a sequence as mentioned in section 2.6 and performed face detection experiments using several image sequences containing near-frontal faces which have variations of position, scale and rotation.

In fig. 2, the results are shown. Each row shows results of Liu's, Verma's, and the proposed method respectively.

As shown in the figure, Liu's frame-wise method made several false-alarms or false-negatives and required a significant computational time. Verma's method and the proposed method had a similar detection and false-alarm rate. But the proposed method had higher locating accuracy when face suddenly varied and required less computational loads. However, we observed some detection failures caused by local extremes of the face likelihood function. The processing time depended largely on the number of iterations and the number of faces. Typically, it was 50ms~1.0sec per frame.

Fig. 2 Face Detection Results using (a) Liu's, (b) Verma's, and (c) the proposed method

4. CONCLUSIONS

In this paper, we proposed a novel face detection method for an image sequence. This paper introduced a new unified Bayesian framework for a sequence and a method to estimate accurate face position, scale, and rotation.

Our proposed framework allows face detection using the whole sequence based on Bayesian decision theory. In the unified framework, with the criteria of face likelihood and similarity between neighboring frames, the most face-like region at each frame can be obtained.

FPGA is a method to obtain accurate face regions based on face probability model. Using the method, face likelihood can be estimated at the situation that given region agrees well with human faces. It increases discriminability between face and non-face, so the number of false alarms or false negatives can be reduced.

In experiments, we have observed that the proposed method is more accurate and faster than other methods.

5. REFERENCES

[1] R.C. Verma, C. Schmid, and K. Mikolajczyk, "Face Detection and Tracking in a Video by Propagating Detection Probabilities," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1215-1228, Oct. 2003.

[2] C. Liu, "A Bayesian Discriminating Feature Method for Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725-740, June 2003.

[3] E. Loutas, I. Pitas, and C. Nikou, "Probabilistic Multiple Face Detection and Tracking Using Entropy Measures," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 128-135, Jan. 2004.

[4] H.I. Kim, S.H. Lee, and N.I. Cho, "Rotation-invariant face detection using angular projections," *Electronics Letters*, vol. 40, no. 12, pp. 726-727, 10th June 2004.