Robust person authentication using dyadic wavelet filters learned by cosine-maximization

Takano, Shigeru
Department of Informatics, Kyushu University

Niijima, Koichi
Department of Informatics, Kyushu University

https://hdl.handle.net/2324/3422

Robust person authentication using dyadic wavelet filters learned by cosine-maximization

Shigeru Takano and Koichi Niijima

December 18, 2006

Abstract

This paper produces a person authentication system using learned lifting dyadic wavelet filters. Our system has the capability of identifying persons whose facial image is changed the size and angle of an original image. The learning algorithm is done by training free parameters in the lifting filters so as to maximize the cosine between a vector whose components are the lifting filters and a vector of facial parts. This problem can be solved fast using Newton's method. Applying the learned filters to a test image, facial parts in the image can be detected. Our detection of facial parts is robust for expansion, shrink and rotation of the image. A person is identified by checking the number of faces detected from video frames. Simulation results show that our person authentication is accurate and fast.

1 Introduction

Person authentication using facial images is a challenging problem in the field of intelligent human computer interaction. This is useful for solving the security problems such as login to a computer, prevention of forged-passport and surveillance camera. There are several studies for identifying persons by using biometrics features of iris and retina [3, 8]. These approaches are high accuracy but require a specific video camera that can analyze iris and retina. So far, many video-based face recognition methods have been developed : color segmentation, template matching, principal component analysis (PCA) [6], support vector machine (SVM) [4], Kohonen's

self organizing map (SOM)[2] and wavelet theory [5]. In these problems, it is very important to develop fast and robust algorithms for learning and identifying facial images. The techniques combining color segmentation with template matching can localize the face position easily. However, it is difficult to identify a person from the localized face. On the other hand, face recognition systems based on PCA, SVM and SOM have enough ability to recognize a person face. These systems require a huge amount of computation to train the features of the face. Wiskott et al. [7] produced the elastic bunch graph matching method based on Gabor wavelet filters. Using a set of components obtained by applying Gabor filters, the method has robustness against the size and rotation of the face. However, since Gabor wavelet transform is a complex valued continuous wavelet, it is time-consuming to compute the set of components.

The goal of this paper is to propose a robust person authentication system based on the recognition of faces exploiting lifting dyadic wavelet filters. Lifting dyadic wavelet filters have a shift-invariant property and lifting term with controllable free parameters. Our learning algorithm trains the free parameters so that the normalized high-pass components of facial parts can be maximized in each of the subbands. These facial parts are both eyes, nose and lips in the training image captured from a USB camera. Concretely, we maximize the cosine between a vector whose components are the lifting dyadic wavelet filters and a vector of facial low-pass components in each of the subbands. This learning algorithm comes down to solve a cosine-maximization problem, which can be solved fast using Newton's method. Then, we input the facial low-pass components obtained by changing the scale and rotation of the training facial parts. We detect the facial parts in the frame by applying the learned filters in each of the subbands to a video frame. If the face of a person is detected in the test frames, then he/she is the target person.

This paper is organized as follows: Chapter 2 introduces lifting dyadic wavelet filters and our learning algorithm. In Chapter 3, we describe detection method of facial parts. In Chapter 4, our person authentication system is presented. Chapter 5 involves experimental results. We close in Chapter 6 with concluding remarks and plans of future work.

2   Learning of lifting dyadic wavelet filters

We describe lifting dyadic wavelet filters which serve as the foundation of our person authentication system. Let $\{h_k^o, g_k^o, \tilde{h}_k^o, \tilde{g}_k^o\}$ be a set of initial dyadic wavelet filters, where h_k^o and g_k^o denote low-pass and high-pass analysis filters, respectively, and $\tilde{h}_k^o$ and $\tilde{g}_k^o$ low-pass and high-pass synthesis filters, respectively. From those filters, we construct a new set of filters $\{h_k, g_k, \tilde{h}_k, \tilde{g}_k\}$ as follows:

$$h_k = h_k^o, \tag{1}$$

$$g_k = g_k^o - \sum_t s_t h_{k-t}^o, \tag{2}$$

$$\tilde{h}_k = \tilde{h}_k^o + \sum_t s_{-t} \tilde{g}_{k-t}^o,$$

$$\tilde{g}_k = \tilde{g}_k^o,$$

where s_t's denote free parameters, which is called a dyadic lifting scheme. We proved that the lifting filters h_k, g_k, $\tilde{h}_k$ and $\tilde{g}_k$ also become dyadic wavelet filters [1]. In this paper, only (1) and (2) are employed.

Let us denote an image by $C_0(i,j)$. The low-pass components of p-th resolution level for $C_0(i,j)$ are constructed successively by the dyadic wavelet transform as

$$C_p(i,j) = \sum_{k,l} h_k h_l C_{p-1}(i + 2^{p-1}k, j + 2^{p-1}l), \quad p = 1, \cdots, P. \tag{3}$$

The free parameters s_t's have to be learned so that the detection of facial parts is robust for expansion, shrink and rotation of a query image. Therefore, we compute high-pass components for an expanded (or shrank) and rotated training image. First, we rotate the coordinates $(i + 2^{p-1}k, j)$ by α_n, expand or shrink the resulting coordinates by r_m, and round off as follows:

$$(i + 2^{p-1}k, j) \rightarrow (i + f_p(kr_m \cos\alpha_n), j + f_p(-kr_m \sin\alpha_n)), \tag{4}$$

where

$$f_p(a) = \begin{cases} \lfloor 2^{p-1}a + 0.5 \rfloor, & a \geq 0, \\ \lfloor 2^{p-1}a - 0.5 \rfloor, & a < 0. \end{cases}$$

Using the new coordinates, we compute

$$C_{p,m,n}^{row}(i,j) = \sum_k h_k C_{p-1}(i + f_p(kr_m \cos \alpha_n), j + f_p(-kr_m \sin \alpha_n)).$$

(5)

Next, rotating the coordinates $(i, j + 2^{p-1}l)$ by α_n, expanding or shrinking the resulting coordinates by r_m, and rounding off as

$$(i, j + 2^{p-1}l) \rightarrow (i + f_p(lr_m \sin \alpha_n), j + f_p(lr_m \cos \alpha_n)),$$

(6)

we compute the high-pass components in horizontal direction

$$D_{p,m,n}(i,j) = \sum_l g_{p,l}^d C_{p,m,n}^{row}(i + f_p(lr_m \sin \alpha_n), j + f(lr_m \cos \alpha_n)),$$

(7)

Here $g_{p,l}^d$ represents the lifting filters (2) in horizontal direction and we write it as

$$g_{p,l}^d = g_l^o - \sum_{t=-L}^{L} s_{p,t}^d h_{l-t}^o.$$

(8)

To calculate the high-pass components in vertical direction, we apply (6) to $C_{p-1}(i,j)$ to get

$$C_{p,m,n}^{col}(i,j) = \sum_l h_l C_{p-1}(i + f_p(kr_m \sin \alpha_n), j + f_p(kr_m \cos \alpha_n)).$$

(9)

Applying (4) to $C_{p,m,n}^{col}(i,j)$, we obtain the high-pass components in vertical direction

$$E_{p,m,n}(i,j) = \sum_l g_{p,l}^e C_{p,m,n}^{col}(i + f_p(lr_m \cos \alpha_n), j + f(-lr_m \sin \alpha_n)).$$

(10)

Here $g_{p,l}^e$ denotes the lifting filter (2) in vertical direction and we write it as

$$g_{p,l}^e = g_l^o - \sum_{t=-L}^{L} s_{p,t}^e h_{l-t}^o.$$

(11)

Now, we determine the parameters $s_{p,t}^d$ and $s_{p,t}^e$ in (8) and (11), respectively, so that the features of facial parts are captured in each decomposition level p. Consider $2(M+1)$ taps of initial low-pass filters $h_{-M}^o, \cdots, h_{M+1}^o$, and three taps of initial high-pass filters, g_0^o, g_1^o, g_2^o. If $M \leq 2$, the index l of $g_{p,l}^d$ and $g_{p,l}^e$ moves from $-L - M - 1$ to $L + M$. We put $N_1 = -L - M - 1$ and $N_2 = L + M$.

For convenience of expression in onward discussion, we define the following four vectors

$$\mathbf{g}_p^d = (g_{p,N_1}^d, \cdots, g_{p,N_2}^d),$$

$$\mathbf{g}_p^e = (g_{p,N_1}^e, \cdots, g_{p,N_2}^e),$$

$$\mathbf{C}_{p,m,n}^{row}(i,j) = \Big(C_{p,m,n}^{row}(i + f_p(N_1 r_m \sin \alpha_n), j + f(N_1 r_m \cos \alpha_n)), \cdots,$$

$$C_{p,m,n}^{row}(i + f_p(N_2 r_m \sin \alpha_n), j + f_p(N_2 r_m \cos \alpha_n))\Big),$$

$$\mathbf{C}_{p,m,n}^{col}(i,j) = \Big(C_{p,m,n}^{col}(i + f_p(N_1 r_m \cos \alpha_n), j + f_p(-N_1 r_m \sin \alpha_n)), \cdots,$$

$$C_{p,m,n}^{col}(i + f_p(N_2 r_m \cos \alpha_n), j + f_p(-N_2 r_m \sin \alpha_n))\Big).$$

Using inner product symbol '·', (7) and (10) can be written in the following forms

$$D_{p,m,n}(i,j) = \mathbf{g}_p^d \cdot \mathbf{C}_{p,m,n}^{row}(i,j), \quad E_{p,m,n}(i,j) = \mathbf{g}_p^d \cdot \mathbf{C}_{p,m,n}^{col}(i,j).$$

Let $\theta_{p,m,n}^d$ and $\theta_{p,m,n}^e$ denote the angles between $\mathbf{g}_p^d$ and $\mathbf{C}_{p,m,n}^{row}(i,j)$, and between $\mathbf{g}_p^e$ and $\mathbf{C}_{p,m,n}^{col}(i,j)$, respectively. Then, the cosine for each of the angles $\theta_{p,m,n}^d$ and $\theta_{p,m,n}^e$ is defined as

$$\cos \theta_{p,m,n}^d = \frac{D_{p,m,n}(i,j)}{|\mathbf{g}_p^d||\mathbf{C}_{p,m,n}^{row}(i,j)|}, \tag{12}$$

$$\cos \theta_{p,m,n}^e = \frac{E_{p,m,n}(i,j)}{|\mathbf{g}_p^e||\mathbf{C}_{p,m,n}^{col}(i,j)|}, \tag{13}$$

where the symbol $|\cdot|$ denotes the Euclidean norm of the vectors.

We learn free parameters $s_{p,t}^d$ and $s_{p,t}^e$ so as to approximate $\mathbf{C}_{p,m,n}^{row}(i,j)/|\mathbf{C}_{p,m,n}^{row}(i,j)|$ by $\mathbf{g}_p^d/|\mathbf{g}_p^d|$, and $\mathbf{C}_{p,m,n}^{col}(i,j)/|\mathbf{C}_{p,m,n}^{col}(i,j)|$ by $\mathbf{g}_p^e/|\mathbf{g}_p|$. This implies that (12) and (13) tend to 1. We notice that the learned filter becomes a low-pass filter though the initial filter is a high-pass filter. These approximations lead us to minimization problems of the functionals

$$J_p^d(\mathbf{s}_p^d) = \sum_{m,n}(D_{p,m,n}(i,j) - |\mathbf{g}_p^d||\mathbf{C}_{p,m,n}^{row}(i,j)|)^2, \tag{14}$$

$$J_p^e(\mathbf{s}_p^e) = \sum_{m,n}(E_{p,m,n}(i,j) - |\mathbf{g}_p^e||\mathbf{C}_{p,m,n}^{col}(i,j)|)^2, \tag{15}$$

where

$$\mathbf{s}_p^d = (s_{p,-L}^d, \cdots, s_{p,L}^d),$$

$$\mathbf{s}_p^e = (s_{p,-L}^e, \cdots, s_{p,L}^e).$$

and (i, j) represents the selected point of facial parts such as eyes, nose and lips. In this paper, we employ Newton's method to seek stationary points of (14) and (15) fast. Newton's method has convergence rate of order two.

Differentiating $J_p^d(\mathbf{s}_p^d)$ and $J_p^e(\mathbf{s}_p^e)$ with respect to each of the free parameters, we obtain the following nonlinear systems of simultaneous equations

$$\frac{\partial J_p^d(\mathbf{s}_p^d)}{\partial s_{p,t}^d} = 0, \qquad t = -L, \cdots, L, \tag{16}$$

$$\frac{\partial J_p^e(\mathbf{s}_p^e)}{\partial s_{p,t}^e} = 0, \qquad t = -L, \cdots, L. \tag{17}$$

We solve (16) and (17) by Newton's method. These equations may have many solutions. So, we find a solution close to zero vector starting Newton iteration from the zero vector.

Applying this learning algorithm to facial parts such as eyes, nose and lips of training faces for a variety of target persons, we learn free parameters in a lifting filter. These learned parameters are memorized in a database together with the training faces.

3   Detection of facial parts

We detect facial parts in a test image using the learned parameters $s_{p,t}^d$ and $s_{p,t}^e$ described in Chapter 2. Let us denote the test image again by $C_0(i, j)$. For $C_0(i, j)$, we compute low-pass components $C_p(i, j), p = 1, \cdots, P$ by (3) and high-pass components $D_p^o(i, j)$ and $E_p^o(i, j)$ in horizontal and vertical directions, respectively, by the formulae

$$D_p^o(i, j) = \sum_{k,l} h_k g_l C_{p-1}(i + 2^{p-1}k, j + 2^{p-1}l), \quad p = 1, \cdots, P,$$

$$E_p^o(i, j) = \sum_{k,l} g_k h_l C_{p-1}(i + 2^{p-1}k, j + 2^{p-1}l), \quad p = 1, \cdots, P.$$

Combining the learned parameters $s_{p,t}^d$ and $s_{p,t}^e$ in the database with $C_p(i, j)$, $D_p^o(i, j)$ and $E_p^o(i, j)$, we compute

$$D_p(i, j) = D_p^o(i, j) - \sum_{t=-L}^{L} s_{p,t}^d C_p(i, j + 2^{p-1}t), \quad p = 1, \cdots, P,$$

$$E_p(i,j) = E_p^o(i,j) - \sum_{t=-L}^{L} s_{p,t}^e C_p(i + 2^{p-1}t, j), \quad p = 1, \cdots, P.$$

To extract the facial parts contained in $C_0(i,j)$, the following quantity is introduced:

$$R(i,j) = \sum_{p=1}^{P} \left((Q_p^d(i,j) - 1)^2 + (Q_p^e(i,j) - 1)^2 \right). \qquad (18)$$

Here $Q_p^d(i,j)$ and $Q_p^e(i,j)$ represent

$$\begin{aligned}
Q_p^d(i,j) &= \frac{D_p(i,j)}{|\mathbf{g}_p^d||\mathbf{C}_p^{row}(i,j)|}, \\
Q_p^e(i,j) &= \frac{E_p(i,j)}{|\mathbf{g}_p^e||\mathbf{C}_p^{col}(i,j)|},
\end{aligned}$$

respectively, where $\mathbf{C}_p^{row}(i,j)$ and $\mathbf{C}_p^{col}(i,j)$, represent (5) and (9), respectively, but now $\alpha_n = 0$ and $r_m = 1$. If the quantity (18) is minimal at the point (i_0, j_0), then $C_0(i_0, j_0)$ provides a facial part we want to extract.

4 Person authentication system

Our person authentication consists of database generation of learned parameters in lifting filters and matching algorithm of facial parts against the database.

4.1 Database of learned parameters

Our database generation is to memorize the learned parameters for each of target persons. We prepare a training image containing a target person's face and select several facial parts such as both eyes, nose and lips from the training image. To learn the parameters robustly, we compute the high-pass components for an expanded (or shrank) and rotated facial part. The lifting parameters are learned so so to maximize the normalized correlation between the lifting filters and the high-pass components of the facial part. A set of the computed parameters is stored in database for each target person.

4.2 Person authentication

In matching process, the lifting filters are constructed by setting the learned parameters in the database. We first apply these lifting filters of a target person to a test image $C_0(i,j)$, and obtain the lifted high-pass components $D_p(i,j)$ and $E_p(i,j)$ in horizontal and vertical directions, respectively, in each decomposition level p. Next, we compute a quantity based on the correlation between the learned lifting filters and these high-pass components. Finding a point (i_0, j_0) in which our quantity is minimal, we get $C_0(i_0, j_0)$ as a facial part. If the test image contains the target person's face, then the facial parts are detected from the test image successfully.

5 Experimental results

In experiments, we used video frames as 12 different training and test images, each of which has 8 bits, 320×240 size. These frames were captured from a USB camera. To reduce the computational effort in learning process, we extract 128×128 images around a nose position given by hand. Figure 1 illustrates the training images including faces of 12 persons. As initial filters, we chose the dyadic
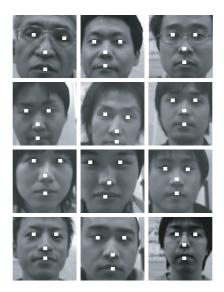

Figure 1: Training images.

wavelet filters whose coefficients are listed in Table 1 and trained the parameters $s_{p,t}^d$ and $s_{p,t}^e$, $p = 1, \cdots, 3$, for facial parts of the training images by learning algorithm described in Chapter 2. To

Table 1: Coefficients of dyadic wavelet filters.

l	h_l^o	g_l^o
-2	0.03125	
-1	0.15625	
0	0.31250	-0.25000
1	0.31250	0.50000
2	0.15625	-0.25000
3	0.03125	

determine the parameters $s_{p,t}^d$ and $s_{p,t}^e$, scale and rotation coefficients r_m and α_n were selected as $(r_0, r_1, r_2, r_3, r_4) = (0.6, 0.8, 1.0, 1.2, 1.4)$ and $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (-\pi/4, -\pi/8, 0, \pi/8, \pi/4)$, respectively.

Applying the learned lifting filters to a video frame, we first detected a nose position of a target person. As the test images for each person, we used 128×128 images which were extracted from the video frames around the first detected position of the target's nose. Then, fast face detection was realized by searching a block area with 10×10 size around nose position detected in $C_0(i,j)$ for both eyes and lips. We computed the sum of $R(i,j)$ in (18) at eyes, nose and lips extracted by applying the detection algorithm described in Chapter 3 to each of the 10 different facial images for each of 12 persons, at the resolution levels $P = 1, 2, 3$. By checking the number of faces detected from these images, we could identify all the persons.

Figure 2 illustrated the face detection results by applying the learned filters for the first author to the 10 test images of the first author and the others. To evaluate performance of our person authentication, we computed the sum of $R(i,j)$ at the facial parts detected in Figure 2 as shown in Table 2. This result indicates that our authentication system have high accuracy for identifying persons.

Simulation was done on a laptop computer with CPU Turion 64 MT34 1.8GHz and memory of 2GB. Computational time of our learning and authentication algorithms is shown in Table 3.

Table 2: Sum of $R(i, j)$ at the detected facial parts.

	Person 1	Person 2	Person 3	Person 4
Frame 0	2.622195	2.617867	2.949306	2.884644
Frame 1	2.635594	2.786780	2.695307	2.835648
Frame 2	2.632414	3.384888	3.258291	2.924920
Frame 3	2.559647	3.124071	2.836267	2.829621
Frame 4	2.626911	2.967742	3.232503	2.844455
Frame 5	2.612055	2.840102	3.208076	3.181805
Frame 6	2.633631	3.590027	3.156549	3.163974
Frame 7	2.612055	2.838915	3.377974	3.497539
Frame 8	2.606328	2.580577	2.913605	2.882148
Frame 9	2.637755	2.685699	3.436957	2.935365

Table 3: Computational time of learning and authentication.

	P=1	P=2	P=3
Learning algorithm	1.1 sec	2.7 sec	4.0 sec
Person authentication	1.9 sec	2.9 sec	3.8 sec

6 Conclusion

We proposed a robust person authentication system for expansion, reduction and rotation for expansion, shrink and rotation of images using the learned lifting dyadic wavelet filters. A main idea of our method is to learn the lifting dyadic wavelet filters by using the cosine of an angle between a vector whose components are lifting high-pass filters and a vector consisting of pixels in the facial parts. The parameters in the lifting filters were determined by using training facial images obtained by changing the size and angle of an original image. Our detection algorithm was performed efficiently as follows: (i) find nose first and then, search around the nose for eyes and lips, (ii) after the second resolution level, search a small region around the detected face at the higher resolution level for facial parts. Our authentication was done by checking the number of faces detected from the video frames. In simulation, our system was carried out in realtime. In the future, we will implement the person expression recognition system on robot vision.
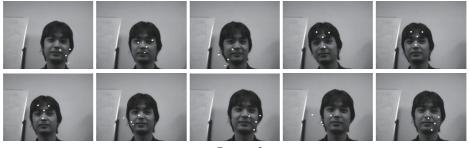
References

[1] T. Abdukirim, K. Niijima and S. Takano, Design of Biorthogonal Wavelet Filters Using Dyadic Lifting Scheme, *Bulletin of Informatics and Cybernetics*, vol.37, pp.123–136, 2005.

[2] G. Burel and D. Carel, Detection and Localization of Faces on Digital Images, *Pattern Recognition Letters*, vol. 15, no. 10, pp. 963–967, 1994.

[3] L. Ma, T. Tan, Y. Wang and D. Zhang, Efficient Iris Recognition by Characterizing Key Local Variations, *IEEE Trans. Image Processing*, vol.13, no. 6, pp. 739–750, 2004.

[4] E. Osuna, R. Freund, and F. Girosi, Training Support Vector Machines: An Application to Face Detection, *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition*, pp. 130–136, 1997.

[5] S. Takano, K. Niijima and T. Abdukirim, Fast Face Detection by Lifting Dyadic Wavelet Filters, *Proc. IEEE Int'l Conf. Image Processing*, pp. 893–896, 2003.

[6] M. Turk and A. Pentland, Eigenfaces for Recognition, *Journal of Cognitive Neuro-science*, vol.3, no.1, pp. 71–86, 1991.

[7] L. Wiskott, J.M. Fellous, N. Krfiger and C. Malsburg, Face Recognition by Elastic Bunch Graph Matching, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.775–779, 1997.

[8] Y. Zhu, T. Tan and Y. Wang, Biometric Personal Identification Based on Iris Patterns, *Proc. Int'l Conf. Pattern Recognition*, vol. II, pp. 805–808, 2000.

Person 1

Person 2

Person 3

Person 4

Figure 2: The detected facial parts.