International Journal of Applied Electromagnetics and Mechanics 15 (2001/2002) 359–364
IOS Press

Application of the Fourier-wavelet transform to moving images in an interview scene

Chieko Kato[a,*], Hisashi Endo[b], Yoshifuru Saito[b] and Kiyoshi Horii[a]
[a]*Shirayuri College, 1-25 Midorigaoka, Chofu-shi, 182-8525, Tokyo, Japan*
Tel.: +81 3 3326 5144 ext.363; Fax: +81 3 3326 7604; E-mail: katoc@bronze.ocn.ne.jp
[b]*Graduate School of Engineering, Hosei University, 3-7-2 Kajino, Koganei, 184-8584 Tokyo, Japan*
Tel.: +81 42 387 6200; Fax: +81 42 387 6200; Email: ysaitoh@ysaitoh.k.hosei.ac.jp

Abstract. A new computer-aided method, based on Saito's Fourier-wavelet theory, is used to extract the facial expressions that appear and disappear in less than one-fifth of a second, i.e. expressions that reveal the true emotions of a human being. The new method consists of two major steps. The first step is the application of the Fourier transform to reduce the spatial phase differences between the reference and test images. The second is the application of the wavelet transform to extract the local discontinuous differences. This method is applied to the moving images in the interview scene. Using this method, the difference between the two facial expressions has been extracted, even though both images differ in their size and reference position.

1. Introduction

The purpose of this study is to capture subtle expressions using a new computer-aided method [1, 2]. In traditional methods, facial expressions were captured by the use of photographic cameras or video cameras [3,4]. In these methods, facial expressions observed by the naked eye and captured in photographs or on videotapes were categorized into six types of emotion – happiness, sorrow, surprise, fear, anger and hatred. However, these methods tended to allow the observer to make a subjective judgment, since the categorization of facial expressions were carried out with the naked eye. When an emotion shows on a face, the change in expression appears and disappears in less than one-fifth of a second [5]. In the traditional methods, it is difficult to capture this change with the naked eye, and the categorization of facial expressions takes a great deal of time; on the other hand, the Fourier-wavelet transform makes it possible to capture the change accurately and greatly reduce the amount of time required for the categorization, with the help of the computer.

This method is based on Saito's Fourier-wavelet theory [1,2]. This method consists of two major stages; the first is to correct positional differences between the reference and test images, and the second is to extract minute differences between the two images. In the present paper, the Fourier-wavelet transform has been applied to the moving images of an interview scene. Nonverbal information especially expressions, as well as words is important in finding out what the interview really means, because these expressions reveal human emotions more precisely than words do.

*Corresponding author: Chieko Kato, Shirayuri College, 1-25 Midorigaoka, Chofu-shi, Tokyo 182-8525, Japan. Tel.: +81 3 3326 5144 ext.363; E-mail: katoc@bronze.ocn.ne.jp.

1383-5416/01/02/$8.00 © 2001/2002 – IOS Press. All rights reserved

2. Fourier-wavelet transform

The Fourier-wavelet transform is a method in which stationary images are used for the extraction of facial expressions. The aim of this study is to extract differences between the reference and test images based on Fourier-wavelet transform. The stationary reference image is obtained by a picture of a passive face of the subject which is taken before the interview. The stationary test images are obtained as follows; first, a video of an interview scene is entered into the computer, and then the motion of the video is stopped every tenth of a second in order to clip stationary images with 128×128 pixels. Excerpts, which are five minutes long in total, from the twenty-minute interview have been used for the extraction of expressions. These excerpts are scenes where the subject felt the happiest.

The changes in facial expressions cannot be captured by the wavelet transform unless the reference and test images are exactly in the same position. However, it is inevitable that positional differences occur where stationary images are clipped. The Fourier transform is used in order to correct the positional differences.

The procedure of the Fourier-wavelet transform based on Saito's Fourier-wavelet theory [1,2] consists of two major stages:

(1) The positional differences are corrected by the use of the Fourier transform.
(2) The changes in the facial expression are extracted by the use of the wavelet transform.

From the view of a discrete matrix operation, a two-dimensional discrete Fourier transform of an image data $\mathbf{F}$ can be given by

$$\mathbf{F}' = \mathbf{C}\mathbf{F}\mathbf{C}^T, \tag{1}$$

where $\mathbf{F}'$ is Fourier spectrum and ^{T} refers to the matrix transpose. The matrix $\mathbf{C}$ is expressed as

$$C = \frac{1}{\sqrt{n}} \begin{pmatrix} e^0 & e^0 & \cdot & e^0 \\ e^0 & e^{i\Delta} & \cdot & e^{i(n-1)\Delta} \\ \cdot & \cdot & \cdot & \cdot \\ e^0 & e^{i(n-1)\Delta} & \cdot & e^{i(n-1)(n-1)\Delta} \end{pmatrix}, e^{i\alpha} = \cos(\alpha) + i\sin(\alpha),\ i = \sqrt{-1}, \tag{2}$$

where n denotes the dimension of the spectra $\mathbf{F}'$, and $\Delta = \Delta x = \Delta y$ denotes the sampling scale of the image along x or y direction.

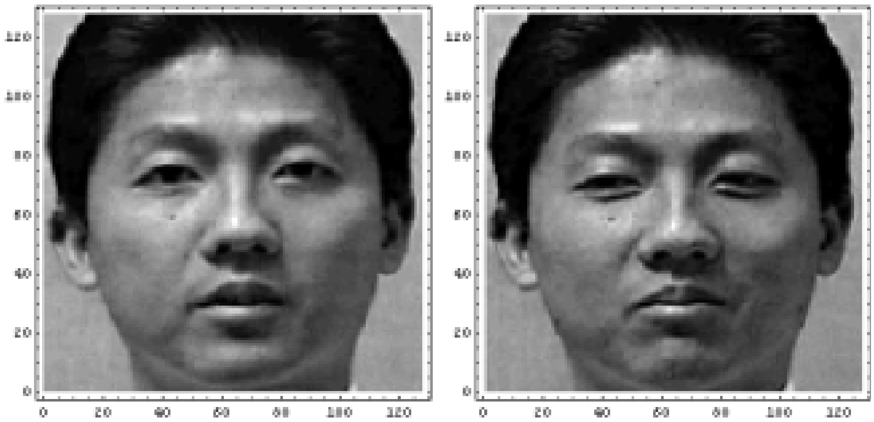
To remove the entire positional difference between the test image $\mathbf{G}$ and reference image $\mathbf{F}$, the compensation filter based on the Fourier transform can be first defined as

$$filter = \frac{\mathbf{G}'}{|\mathbf{G}'|} - \frac{\mathbf{F}'}{|\mathbf{F}'|}. \tag{3}$$

where $||$ is the magnitude of the spectrum. Then by convolving this compensation filter with spectrum $\mathbf{G}'$ of the test image, a different region of spectrum between the reference and test images is retained. To reproduce a different image $\mathbf{D}_f$, describe the spatial frequency difference between the reference and test images, the inverse Fourier transform is carried out by

$$\mathbf{D}_f = (C^T)^*(filter \times \mathbf{G}')(C)^*. \tag{4}$$

where the superscript $*$ refers to the complex conjugate.



(a) Reference image **(b) Test image at t = 0 min**

Fig. 1. Images of a human face.

As for the second step of Fourier-wavelet transform, the wavelet multi-resolution analysis is adopted for the extracting of local or small differences in the grouped different frequency domain. The two-dimensional discrete wavelet transform of the reproduced different image $\mathbf{D}_f$, can be defined as

$$\mathbf{S} = \mathbf{W}\mathbf{D}_f\mathbf{W}^T, \tag{5}$$

where $\mathbf{S}$ and $\mathbf{W}$ represent the wavelet spectrum of $\mathbf{D}_f$ and the analyzing wavelet matrix, respectively. A cascade algorithm of an orthogonal wavelet basis function is usually obtained by analyzing the wavelet matrix. Since the discrete wavelet transform is an orthogonal transform, the inverse wavelet transform is simply carried out by

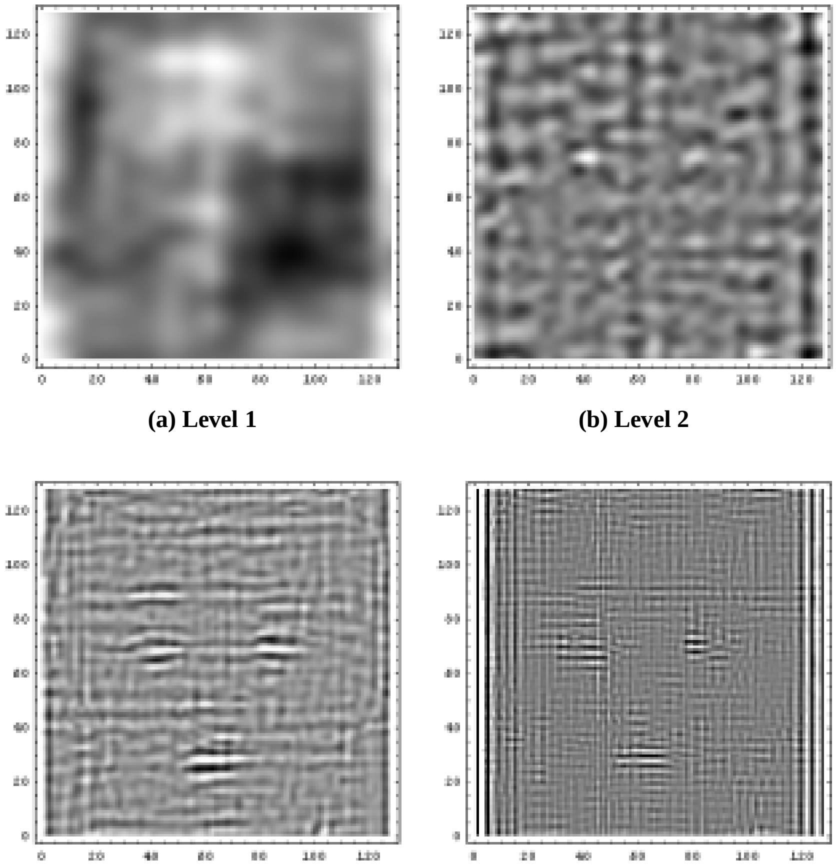
$$\mathbf{D}_f = \mathbf{W}^T\mathbf{S}\mathbf{W}. \tag{6}$$

It is a well-known fact that the discrete wavelet transform is a linear transform and the image can be decomposed into grouped frequency components of the images. This decomposition method is called the wavelet multi-resolution analysis. Hence, Eq. (6) can be rewritten by

$$\mathbf{D}_f = \mathbf{W}^T\mathbf{S}_1\mathbf{W} + \mathbf{W}^T\mathbf{S}_2\mathbf{W} + \cdots + \mathbf{W}^T\mathbf{S}_n\mathbf{W}. \tag{7}$$

On the right side of the equation above Eq. (7), the first term $\mathbf{W}^T\mathbf{S}_1\mathbf{W}$ and the last term $\mathbf{W}^T\mathbf{S}_n\mathbf{W}$, $i = 1, 2 \ldots, n$ represent the image components at level 1 (the lowest frequency) and level n (the highest frequency).

To demonstrate the feature of the Fourier-wavelet transform, two images of a human face expressing the different facial expressions in Fig. 1 are analyzed. The left image of Fig. 1(a) is a passive face (reference image) and the right image of Fig. 1(b) is a smiling face (test image). In this study, the Coifman basis function with orders 30 is used, which not only has smoothness, but also detailed localization characteristics in the frequency domain. The difference image is decomposed into four wavelet levels by the Fourier-wavelet transform. The difference image components are shown in Fig. 2. It is obvious that the difference image components exhibit the local differences from low to high spatial frequencies as the level is changed from 1 to 4. The largest difference is displayed as black and the smallest as white. The local differences between the passive and smiling faces are easily observed under the left eye



(a) Level 1 (b) Level 2

(c) Level 3 (d) Level 4

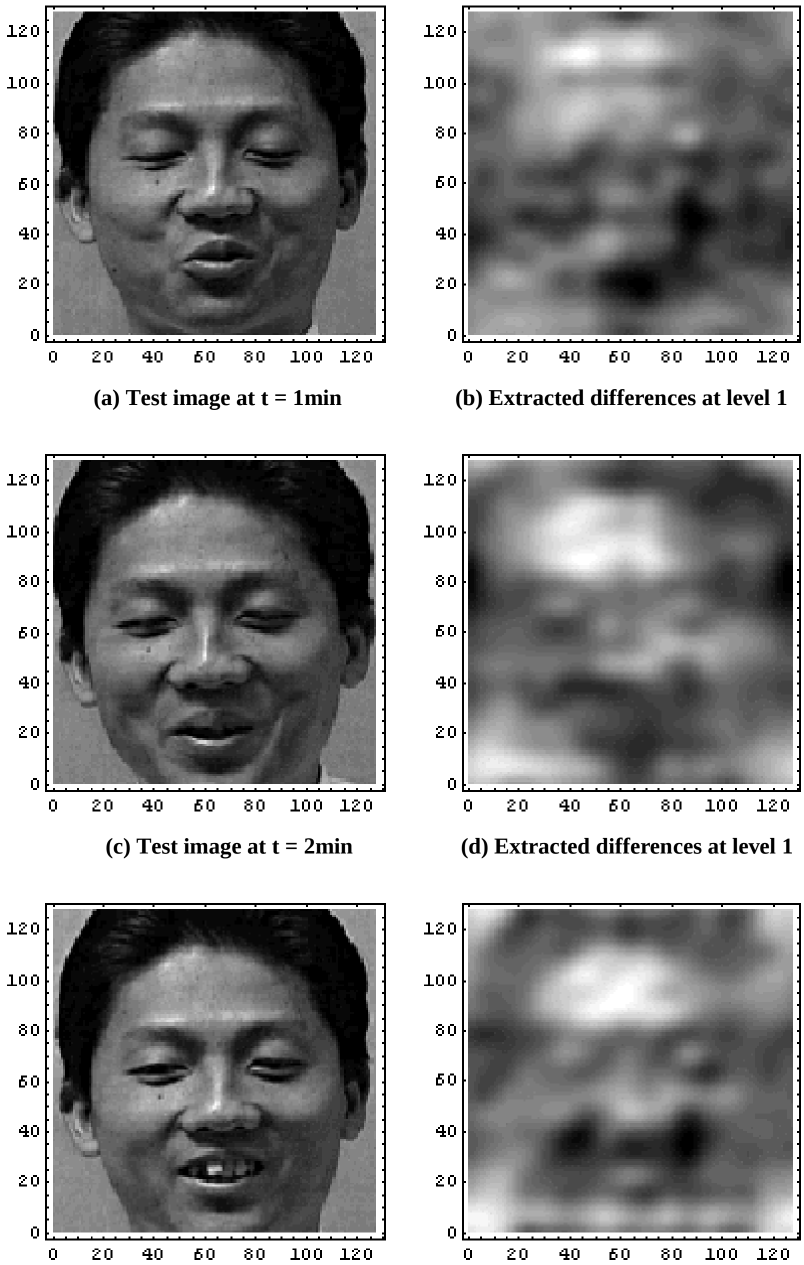
Fig. 2. Results of Fourier-wavelet transform.

and on the left cheek at level 1. This means that the Fourier-wavelet transform can offset the positional differences and can extract much detailed information on the local differences. In the following section, the difference image component of level 1 is only used to analyze the facial expression difference, since it exhibits more evident local differences than other levels.

3. Results and discussion

As mentioned previously, a twenty-minute interview was conducted in which facial expressions linked to happiness were observed. From the interview five minutes of happiness was collected and separated into one tenth of a second clips. The expressions of happiness can be located in the following areas of the face in previous research [4] using the naked eye:

- Wrinkles are seen under the eyes.
- Both ends of the lips are horizontally pulled or raised toward the ears.
- The cheeks are raised.
- The mouth is sometimes so open that the teeth show.



(a) Test image at t = 1min **(b) Extracted differences at level 1**

(c) Test image at t = 2min **(d) Extracted differences at level 1**

(e) Test image at t = 3min **(f) Extracted differences at level 1**

Fig. 3. Application of the Fourier-wavelets transform to moving in an interview scene.

– Wrinkles are seen running downward from both sides of the nose beyond both ends of the lips.

With the use of the naked eye previous researchers could observe changes in facial expression showing happiness without a clear measurement of time.

The current research, in which the Fourier-wavelet transform was applied, uncovered three new discoveries. First, more detailed changes in facial expressions associated with happiness were found. Second, the use of the computer makes it possible to accurately capture the changes in the expression in shorter intervals. Third, with the use of computer technologies the process of observing these changes was more objective than in previous research.

A sample of the findings from the current research is provided to illustrate in detail the benefits of the Fourier-wavelet transform. The differences extracted by the Fourier-wavelet transform represent the subject's expressions of happiness in accordance with the conventional categorized features. Figures 1(b), 3(a), 3(c) and 3(e) show the test images of different facial expressions from $t = 0$ to 3 min. To extract differences between the reference image (Fig. 1(a)) and the test images of each time, the Fourier-wavelet transform is used. The difference image components of level 1 at different instances are shown in Figs 2(a), 3(b), 3(d) and 3(f). The biggest differences are observed under the left eye and on the left cheek in Fig. 2(a). The second biggest differences in this result are observed under the right eye and on the wrinkles running from both sides of the nose to both ends of the lips. In Fig. 3(b), the biggest differences are observed under the eyes, on the cheeks, and on the chin. The second biggest differences are observed on the wrinkle running from left side of the nose to left end of the left lip. In Fig. 3(d), the biggest differences are observed on the mouth, on the cheeks, and on the chin. Further, in this result, the second biggest differences are observed under the right eye and on the wrinkle running from right side of the nose to right end of the right lip. In Fig. 3(f), the biggest differences are observed on the cheeks. The second biggest differences are observed under the eyes, on the eyebrows, on the mouth, and on the wrinkles running from both sides of the nose to both ends of the lips.

As can be seen from the samples mentioned above, the Fourier-wavelet transform shows for the first time that differences in facial expression appear on the chin when the subject feels happy. The Fourier-wavelet transform enables the gradation of differences between the reference and test images. Furthermore, areas of blackness found on the face highlighted the expressions of happiness more clearly than previous research could.

4. Conclusions

The Fourier-wavelet transform allows the researcher to study changes on the facial expressions in more detail and more precisely. Moreover, this method of testing enables the researcher to receive information about a new area of the face that expresses happiness. However, the Fourier-wavelet transform is in the very early stages of development and trial therefore, both further improvement and practical experimentation are required. In the future, new computer programs will make the process of evaluating expressions of happiness easier. For example, the process of judging the degrees of difference in facial expressions will be performed efficiently by the computer. Overall, the experiment of combining the Fourier transform and wavelet transform by the use of a computer has been a success.

References

[1] Y. Saito, *Introduction to Image Processing by Mathematica,* in Japanese, Asakura Press, 1998.
[2] Y. Saito, *Wavelet Transform by Mathematica*, in Japanese, Asakura Press, 1998.
[3] H. Ernst, *Evaluation of Facial Musculature and Facial Expression*, Johns Hopkins Press, 1931.
[4] P. Ekman and W.V. Friesen translated by T. Kudo, *Unmasking the Face,* in Japanese, Seishin Press, 1982.
[5] T. Kudo, *Psychological Analysis of Gestures and Expressions*, in Japanese, Fukumura Publishing Company, 1999.