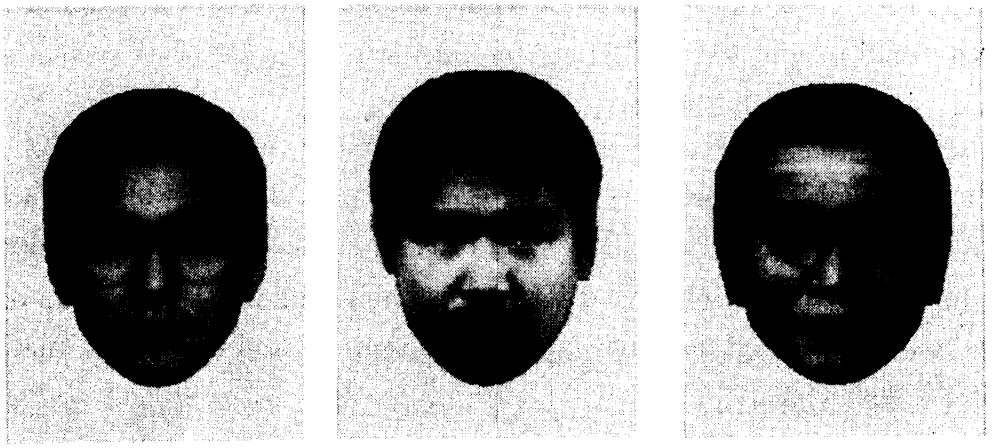
Face, Expression and Communication

- From Facial Information Science to Human Communications Engineering -

Hiroshi Harashima
The University of Tokyo

Each the three pictures in Fig. 1 is a face taken from the average of about ten people from three different occupations. Can you guess these three occupations?

No individual in the world looks exactly the same as any of these faces because these are average faces. They are, so to speak, virtual people. However, by taking the average of various occupations it is possible to bring out the features in these faces. Later in this paper the three occupations will be revealed.



a) b) c)

Fig. 1

1 What is a face ?

Since birth each of us has maintained a relationship with our own face. When the face you see in the mirror in the morning appears to be in a good mood, the day somehow turns out to be a good one. Whether you like it or not, your face is your lifelong companion.

However, our faces are also ourselves. They are a source of pride and a cause of psychological complexes. The Japanese expression "To have mud slung at one's face (kao ni doro wo nuru)" means that your dignity itself has been injured.

Furthermore, your face is a vehicle for communication. We frighten people by expressing anger, express friendliness with a smile. For most people, the impression given by their face has a larger impact than most other factors.

Faces also act as windows of the soul. We can tell somewhat about a person's emotions and feelings by their face and its expression. Conversely, politicians endeavor to hide their feelings with a forced smile.

2 "Faces" as an interdisciplinary field

It may seem that the author of this paper is a psychologist, researching faces. Actually, I belong to Department of Electronic Engineering, and my speciality is a visual telecommunications, with particular emphasis on visual telephones. However, I am not referring to common visual telephones, but rather am dreaming of the following system employing computer graphics for the ultimate in communications.

Because it is a visual telephone, it is only natural that we will be showing our face to other people when we use a visual telephone. However, for example, many people would feel reluctant to use a visual telephone early in the morning without their makeup on. In this case, if we could send a picture of our face with makeup on, and manipulate that image on the visual TV telephone...

To do this we prepare a facial picture attached to a facial frame model in the receiver's

computer, and we move this on the basis of movement information from the sender. This is the basic concept of intelligent image coding which I have been researching.

By using this technology, it is no longer just a dream to put a face on a computer. For example, let us suppose that a robot with a human face on the screen of the monitor appeared. You could talk with the robot just as if you were talking on a visual telephone. If this were the case, even a complete beginner would be able to interact with a computer quite freely. Furthermore, if the computer's internal state could be viewed by looking at the expression of the robot, it would literally be possible to have instantaneous communicate.

Researching faces is not solely the domain of engineering. A great many fields are involved. Anthropology, psychology, physiology, cosmetology, physiognomy, cosmetic surgery, orthodontics, forensic medicine, theater, the arts,..... Each field has a different point of view, but more or less, each deals with the face. In this respect, the face is the quintessential interdisciplinary field.

However, there is no quantitative and objective methodology to deal with the research of the human face or that of human expressions. Therefore, conventional research of this nature stopped at simple qualitative descriptions. As mentioned above communications was my original field of research, however, recently my attention has been drawn to this interdisciplinary field as a research tool for the face.

Below I would like to introduce an outline of the analysis-synthesis system of facial appearances and expressions, developed from the abovementioned standpoint.

3 Synthesizing facial models

In the analysis-synthesis system of facial images, we first prepare a face and head. How we use this to generate an image is the issue.

Here we represent facial features, or an arrangement of the head, the shape of the face, eyes, nose and mouth with a 3-D structure model as in Fig.2. This corresponds to a wire-frame model in the field of computer graphics. It is fine to think of this structure as a facial frame.

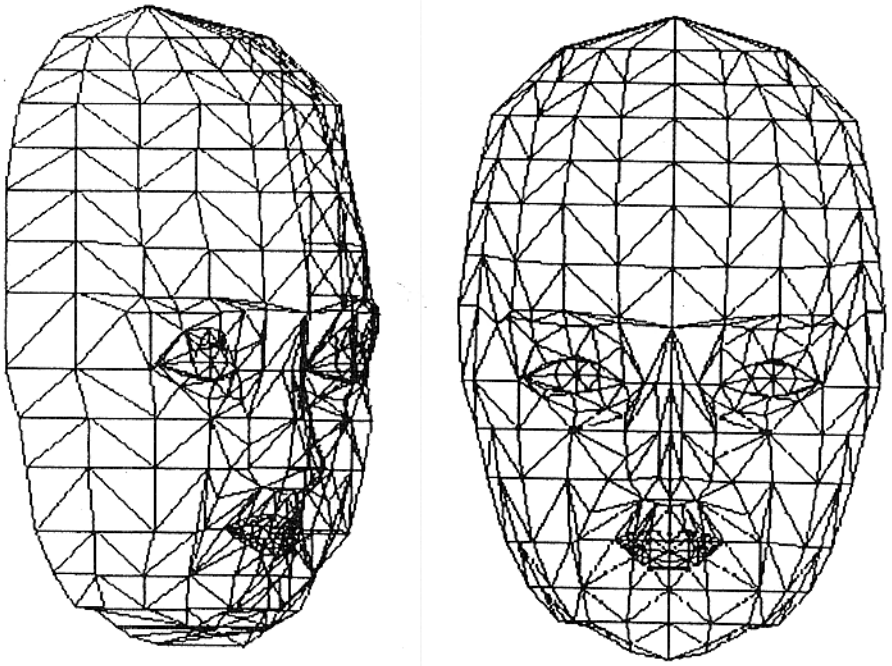


Fig.2

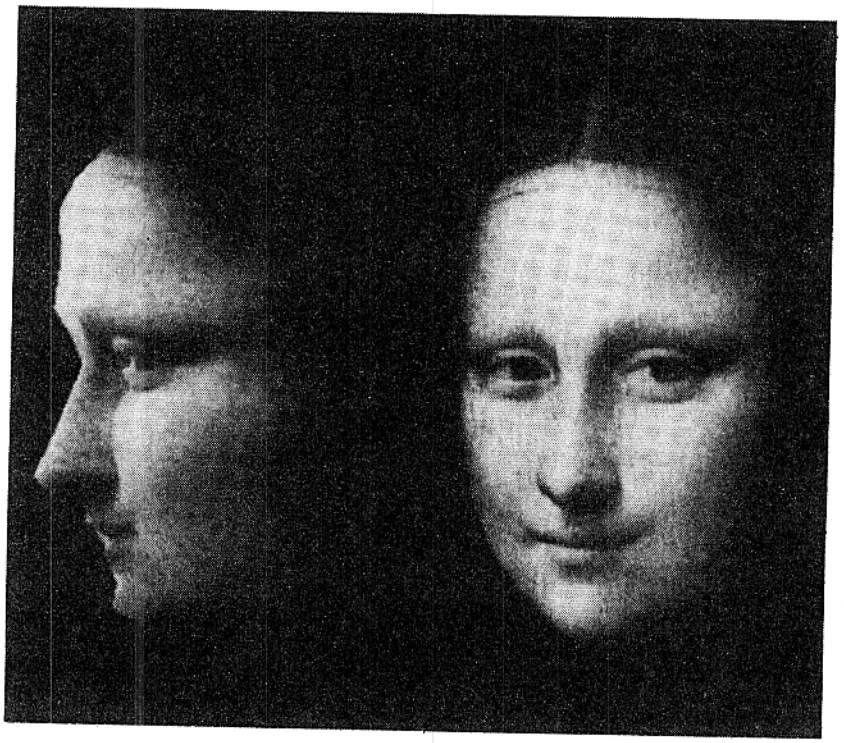


Fig.3

For example, if we were to project a picture of a specific individual's face onto this facial frame and skin to the surface, we can generate a facial image with texture.

Once the facial frame is prepared in this way we can synthesize a facial image which can be seen from any direction. Fig.3 is a synthesized example of Mona Lisa's profile.

4 Synthesizing facial expressions

The next issue is to add facial expressions to the facial frame, namely the facial model. In general, facial expressions are made by the movement of specific points around the eyes and the mouth caused by expansion and contraction of facial muscles. Therefore, we can synthesize a facial image with facial expressions by parameterizing this movement and transforming the structural model.

FACS (Facial Action Coding System)[3] by Ekman and Friesen is known as an attempt to describe facial expressions by the movement of specific facial points. In the description, they analyze the 44 basic movements called AUs (Action Units) coincident with changes in human facial movement. According to them, every facial expression can be expressed by combinations of these AUs.

Recreating this on a computer, we can efficiently synthesize any facial expression by transforming the 3-D structure model through the strength of each AU, with the skin moving at the same time.

In this current research, 34 AUs have been recreated [4]. Synthesized examples of facial expression using AUs are shown in shown Fig.4.

5 Synthesizing facial appearances

Adding facial expressions by FACS was based on geometrical transformations of structural models. However, facial features are not always only geometrical shapes. Facial shadows or skin conditions are important factors in characterizing faces. It is important to grasp these

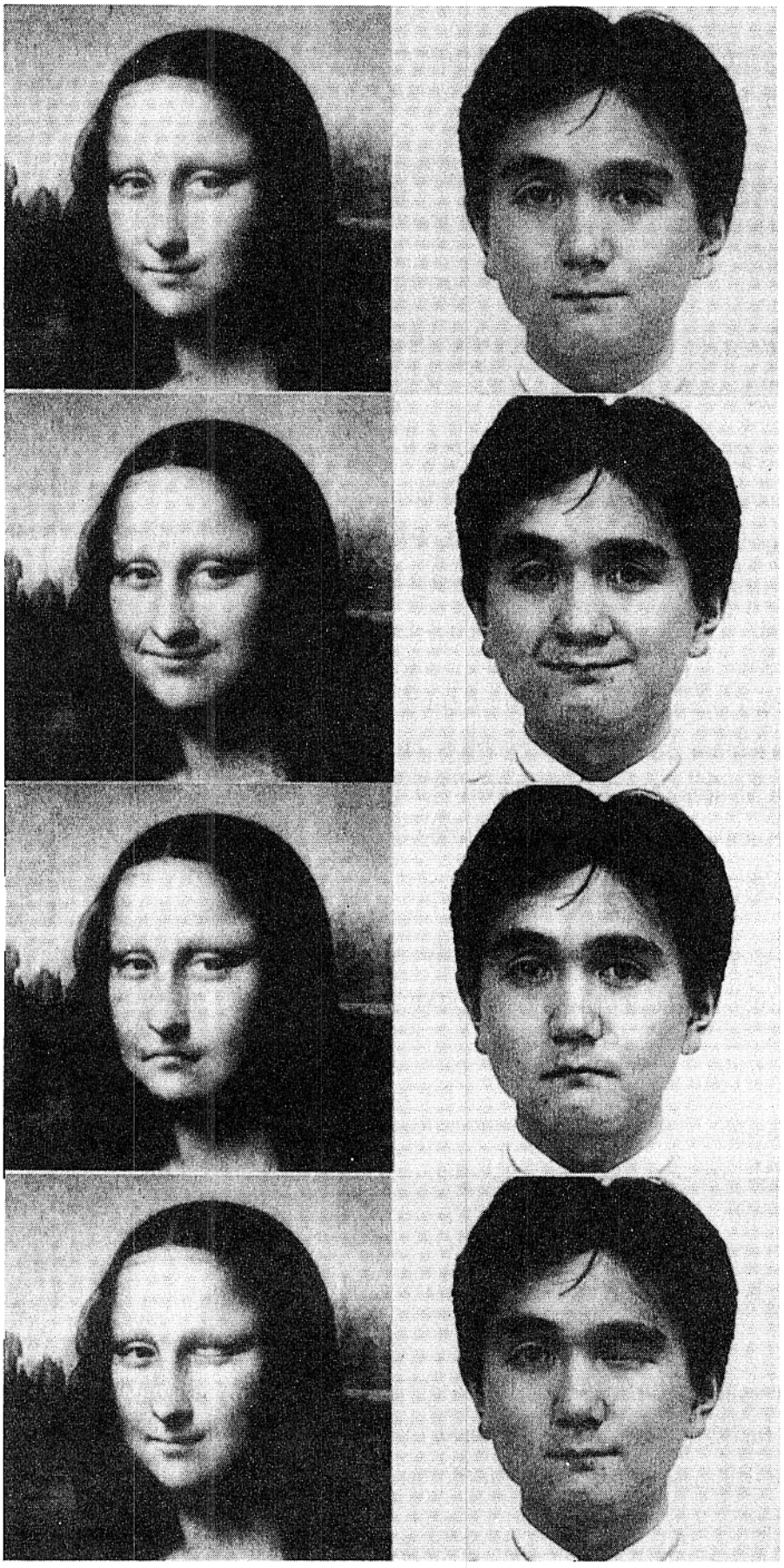


Fig.4.

as unified appearance patterns for the face.

How can we parameterize these facial appearance patterns? As one of the answers to this, in this current research, We are attempting to parameterize by the basic expansion of facial images as follows.[5]

In the field of signal analysis, the signal waveform $f(t)$ can be expressed as

$$f(t) = \sum_k C_k \Phi_k(t) \tag{1}$$

where $\Phi_k(t)$ is the basis of the waveform. If we make the basis a sinusoidal wave, it becomes well-known Fourier analysis.

For facial images, the basis itself is also considered as a facial image. We prepare several different basic facial images, and it is possible to express a new facial image by weight combination. However, simply figuring out the weighted sum on the pixel levels will create an unfocused picture. So the following measures need to be taken;

1. First, by calculating the weighted sum against the coordinates of each facial point (for example, the coordinates of the node of each structural model) we can get the desired facial shape.
2. Next, we geometrically transform the texture pattern of the basic facial image, and make that equal to the facial shape we got in 1.
3. We calculate the weighted sum for each pixel value of the facial image basis of the same shape, and synthesize the facial image.

In this way we are able to get a natural facial image by joining together the facial image bases.

On the other hand, when the facial image is given, we can get an approximation of the facial image by the weighted sum of the facial image bases, and get the parameter as well (weight: corresponds to equation (1),C_k). In that case, having the facial image bases orthogonal

to each other is desirable. This orthogonalization can be achieved by the technique of principal-component analysis (or the Karhunen-Loeve expansion)[5].

☆ ☆ ☆

The first three facial pictures in Fig.1 were synthesized by this technique, They are average male faces from specific occupations [6]. Each occupation is as follows:

a) Average faces of 13 bank clerks

b) Average faces of 11 professional wrestlers

c) Average faces of 10 faction-leader level Liberal Democratic Party members

Professional wrestlers have a narrow gap between their eyebrows and eyes, giving the impression of threatening their opponents. On the contrary, bank clerks have faces which seem to be inviting. The average politician face has narrow eyes. Age also may be a factor, but the politicians seem not to want their expressions read.

6 Expressing emotions and impressions

It is no longer a dream for us to employ analysis-synthesis on faces and facial expressions to analyze human emotions, or on the contrary, to synthesize facial with emotion.

For example, in the field of psychology, there are six basic human emotions: surprise, fear, disgust, anger, happiness and sadness. Therefore, we prepare these six expressions as facial image bases, get a weighted synthesis, and can synthesize facial images with intermediate facial expressions. On the other hand, by using parameterization technology for facial expressions mentioned in 4. and 5., we can get various facial expressions from facial images and plot those points in space. For example, by using FACS, 44 dimensional space, which equals the number of AUs, can be created. The facial expressions can be positioned as coordinates to the strength of AUs. When synthesizing facial expressions from facial image bases, the weight itself will be the coordinate points in the space.

If we can discover a correspondence between these facial expression spaces, or faces with facial appearance spaces which have bases mentioned in 5. and psychological knowledge [7], we may come to understand the relation between emotions and facial expressions, personality and facial appearance. Now we are at the preparatory investigation stage, searching for the relation between emotional words like "bitter experience" and facial expression space, and the relationship between facial impression words like "kind face" and facial appearance space.

7 Facial information science, then toward human communications engineering

Above I have briefly covered the outline of the analysis-synthesis system for facial image and expressions. Our present goal is to develop this study into Facial Information Science and Kansei Communication Science. The Grant-in-Aid for Scientific Research on Priority Areas Project "Information science and psychological research of Kansei information processing" was started by the Ministry of Education last year, with facial research as one of the key elements.

However, as I mentioned above, my speciality is communications. From that view of point, I am dreaming of the creation of a new engineering field, "Human Communications Engineering", by fusing Kansei communications science and conventional electric communication technology.[8]

Originally electric communication developed as information transmission technology between telephone terminals. However, from the point of view of communications, the original terminal was not a telephone but humans themselves. The goal of Human Communications Engineering to re-grasp communication technology as a communication support technology for humans, not simply as an information transmission technology.

For example, if the party you are communicating with happens to be far away, conventional telecommunication technology is required to fill the communication gap. However, even people in close proximity have communication gaps. There might be a gap for people who are deaf, blind, or those with language gaps. There may even be communication gaps

between parents and children. We hope that communication media can fill these kinds of gaps. To do this we need quite a lot of technological help. Recently multimedia and virtual reality have become popular topics. I believe these are within the fields of Human Communications Engineering.

In any case, future media will be expected not only to communicate exactly, but rather Kansei communication media, which can improve the unification among people is what is needed. Japanese people are considered a race who are not good at Kansei communication through facial expressions and gestures. I hope that in some way I can contribute, through research on Human Communications Engineering, to achieving a society rich in Kansei communication.

References

[1] Hiroshi Harashima: "Intelligent image coding and communication", J. Inst. of Television Engineers of Japan, 42, 6, pp.519-525 (June 1988)

[2] Hiroshi Harashima: "Intelligent image coding and video-robot technology", J. of the Society of Instrument and Control Engineering, 30, 6, pp.478-484 (June 1991)

[3] P.Ekman, and W.V.Friesen: "Facial Action Coding System", Consulting Psychologists Press (1977)

[4] C.S.Choi, H.Harashima and T.Takebe: "3-dimensional model-based description and synthesis of facial expressions", Trans. IEICE, J73-A, 7, pp.1270-1280 (July 1990)

[5] H.Harashima, T.Okazaki, C.S.Choi and T.Takebe: "Principal-component analyis of facial images and its applications", IEICE Technical Report, HC90-28 (Jan.1991)

[6] "Science with a good face and a bad face", Quoak, 128, pp.62-69 (Feb.1993)

[7] Yasuyuki Fukui: "Psychology of Emotions", Kawashima Shoten (1990)

[8] Hiroshi Harashima: "Towards human oriented communication age", IEICE Technical Report, HC90-1 (April 1990)