Coding Facial Expressions with Gabor Wavelets

Michael Lyons and Shigeru Akamatsu
ATR Human Information Processing Research Laboratory
2-2 Hikaridai, Seika-cho
Soraku-gun, Kyoto 619-02, Japan
mlyons@hip.atr.co.jp

Miyuki Kamachi and Jiro Gyoba
Psychology Department, Kyushu University

Abstract

A method for extracting information about facial expressions from images is presented. Facial expression images are coded using a multi-orientation, multi-resolution set of Gabor filters which are topographically ordered and aligned approximately with the face. The similarity space derived from this representation is compared with one derived from semantic ratings of the images by human observers. The results show that it is possible to construct a facial expression classifier with Gabor coding of the facial images as the input stage. The Gabor representation shows a significant degree of psychological plausibility, a design feature which may be important for human-computer interfaces.

1. Introduction

In contrast with current human-computer interaction, face-to-face human communication uses a variety of modes, including facial gestures and expressions. It would be desirable to make use of more natural communication modes in human-computer interaction and ultimately computer facilitated human interaction. The automatic recognition of natural facial expressions is a necessary step towards this goal.

Several classes of perceptual cue to emotional state are displayed by the face: relative displacements of features (opening the mouth); quasi-textural changes in the skin surface (furrowing the brow); changes in skin hue (blushing); and the time course of these signals. This paper considers feature displacement and quasi-textural cues. Motion is considered implicitly through comparison of images. Color is not considered here.

The general framework for representing facial images used here is based on topographically ordered, spatially localized filters to represent pattern in the image. The filters consist of a multi-resolution, multi-orientation bank of Gabor wavelet functions. A similar representation is used in the automatic face recognition system developed by the von der Malsburg group [7].

Previous work on automatic facial expression processing includes studies using representations based on: optical flow estimation from image sequences [10, 15, 1]; principal components analysis of single images [2, 1]; and physically-based models [5]. This paper describes the first study that uses Gabor wavelets to code facial expressions. Our findings indicate that it is possible to build a automatic facial expression recognition system based on a Gabor wavelet code which has a significant level of psychological plausibility. This conclusion is supported by the recent work of Zhang et al. [16] demonstrating expression classification using Gabor coding and a multi-layer perceptron.

2. Gabor coding of facial expressions

To extract information about facial expression, each 256 by 256 pixel image, I, was convolved with a multiple spatial resolution, multiple orientation set of Gabor filters (Fig. 1), $G_{\vec{k},+}$ and $G_{\vec{k},-}$. The sign subscript indicates filters of even and odd phase, while $\vec{k}$, the filter wave-vector, determines the spatial frequency and orientation tuning of the filter. A description of the complex-valued two dimensional Gabor transform is given by Daugman [3]. Responses of the filters to the image were combined into a vector, $\mathbf{R}$, with components given by:

$$R_{\vec{k},\pm}(\vec{r}_0) = \int G_{\vec{k},\pm}(\vec{r}_0, \vec{r}) I(\vec{r}) d\vec{r},$$

where,

$$G_{\vec{k},+}(\vec{r}) = \frac{k^2}{\sigma^2} e^{-k^2 \|\vec{r}-\vec{r}_0\|^2/2\sigma^2} cos(\vec{k} \cdot (\vec{r} - \vec{r}_0)) - e^{-\sigma^2/2}),$$

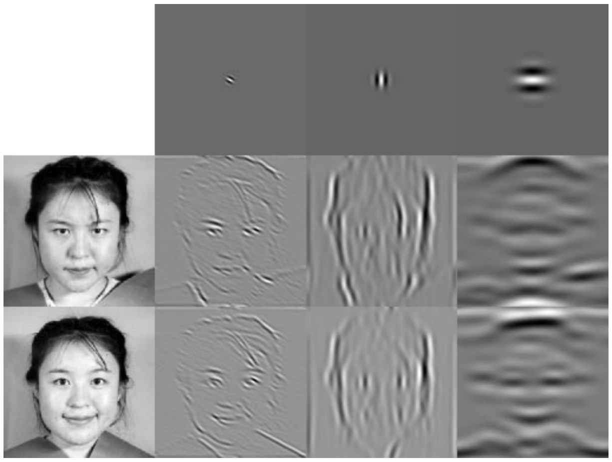

Figure 1. Examples of Gabor filter responses to two facial expression images for three of the filters used.

$$G_{\vec{k},-}(\vec{r}) = \frac{k^2}{\sigma^2} e^{-k^2\|\vec{r}-\vec{r}_0\|^2/2\sigma^2} sin(\vec{k}\cdot(\vec{r}-\vec{r}_0)).$$

The integral of the cosine Gabor filter, $e^{-\sigma^2/2}$, was subtracted from the filter to render it insensitive to the absolute level of illumination. The sine filter does not depend on the absolute illumination level. Three spatial frequencies were used with wavenumbers $k = \{\frac{\pi}{2}, \frac{\pi}{4}, \frac{\pi}{8}\}$ measured in inverse pixels. The highest frequency is set at half the Nyquist sampling frequency, with frequency levels spaced at octaves; $\sigma = \pi$ was used in all calculations, giving a filter bandwidth of about an octave, independent of the frequency level. Six wavevector orientations were used, with angles equally spaced at intervals of $\frac{\pi}{6}$ from 0 to π.

The components of the Gabor vector, $R_{\vec{k}}$, are defined as the amplitude of the combined even and odd filter responses $R_{\vec{k}} = \sqrt{R^2_{\vec{k},+} + R^2_{\vec{k},-}}$. The response amplitude is less sensitive to position changes than are the linear filter responses.

To study the similarity space of Gabor coded facial images, responses of filters having the same spatial frequency and orientation preference were compared at corresponding points in the two facial images. The normalized dot product was used to quantify the similarity of two Gabor response vectors. The similarity of two facial images was calculated as the average of the Gabor vector similarity over all corresponding facial points. Since Gabor vectors at neighboring pixels are highly correlated and redundant, it is sufficient to calculate the average on a sparse grid covering the face (Fig. 2). This similarity measure is used in the automatic face recognition system developed by the von der Malsburg group [7]. The filter parameters used here differ from those used in that work. Automatic systems for scaling the face and registering a graph approximately with the features of the face have been demonstrated previously [7, 14]. In this study, for higher precision, facial graphs were positioned manually on images of a standard scale.

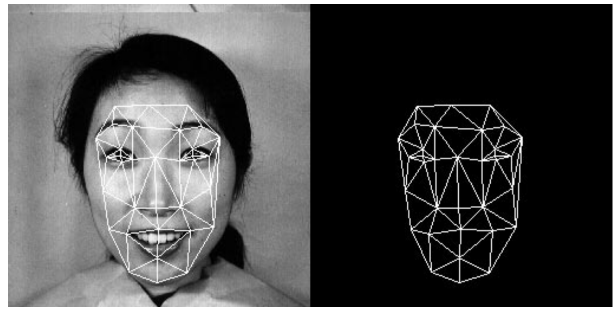

Figure 2. The 34 node grid used to represent facial geometry.

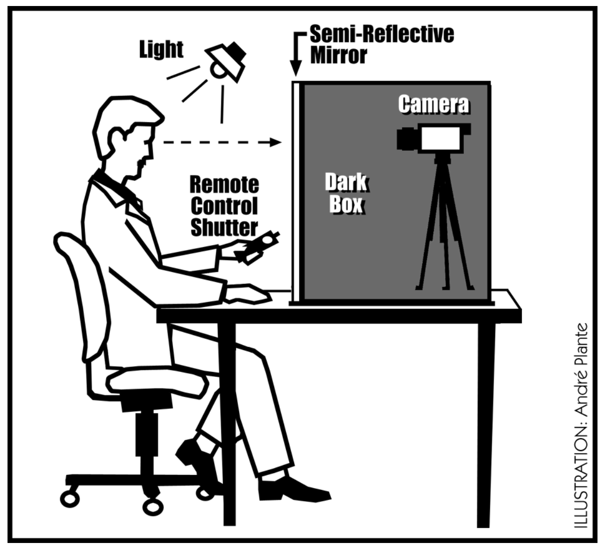

Figure 3. Apparatus used to photograph facial expressions.

3. Facial expression database

A database of facial expression images was collected. Ten expressors posed 3 or 4 examples of each of the six basic facial expressions (happiness, sadness, surprise, anger, disgust, fear) [4] and a neutral face for a total of 219 images of facial expressions. For simplicity of experimental design only Japanese female expressors <and subjects were employed. Fig. 3 shows the apparatus used to photograph the expressors. Each expressor took pictures of her-

Figure 4. Examples of images from the facial expression database.

self while looking through a semi-reflective plastic sheet towards the camera. Hair was tied away from the face to expose all expressive zones of the face. Tungsten lights were positioned to create even illumination on the face. A box enclosed the region between camera and plastic sheet to reduce back-reflection. The images were printed in monochrome and digitized using a flatbed scanner. Sample images are shown in Fig. 4

4. Semantic rating of facial expression images.

Experimental subjects rated pictures for degree of each component basic expression on a five point scale. A total of 92 Japanese female undergraduates took part in the study. The subject pool was divided into four groups: 1.A, 1.B, 2.A, and 2.B. Group 1.A (31 subjects) rated 108 pictures on six basic facial expressions (happiness, sadness, surprise, anger, disgust and fear). Group 1.B (31 subjects) rated the complementary set of 111 pictures (out of the 219 total) on the six basic expression. Both Group 1.A and 1.B saw images of all seven expression categories (including fear images). Group 2.A (15 subjects) rated 94 pictures on five of the six basic facial expressions (fear was excluded). Group 2.B (15 subjects) rated a different set of 93 images on the five basic facial expressions (fear excluded). The images presented to Group 2.A and 2.B excluded fear expressions. Each image was thus labelled with a 5 or 6 component vector with ratings averaged over all subjects. Similarities between these semantic vectors were calculated using the Euclidean distance.

5. Results

Facial expression image similarity computed using the Gabor coding and semantic similarity computed from human ratings were compared by rank correlation. It is convenient to compare similarity spaces rather than categorization performance as this avoids the problem that posed expressions are not always pure examples of a single expression category.

Expressor Initials	Gabor	Geometry
KA	0.593	0.467
KL	0.465	0.472
KM	0.616	0.527
KR	0.636	0.368
MK	0.472	0.287
NA	0.725	0.358
NM	0.368	0.099
TM	0.423	0.282
UY	0.648	0.074
YM	0.538	0.455
Average	0.568	0.366

Table 1. Rank correlation between model and semantic rating similarities.

Geometric similarity was also rank correlated with the semantic ratings similarity values, as a control. The distance of each grid point (Fig. 2) from the point at the nose tip formed the components of a 33 dimensional shape vector. Dissimilarity between two grid configurations were calculated using the euclidean distance.

For the experiments in which all facial expressions were included (i.e. comparison with data from subject groups 1.A and 1.B) the rank correlation between Gabol model and human data ranged from 0.42 (expressor TM) to 0.725 (expressor NA) with an average value of 0.568. For the geometry based control, rank correlation between model and data ranged from 0.074 (expressor UY) to 0.527 (expressor KM) with an average value of 0.366. Correlation results for all expressors are listed in Table 1. With fear stimuli and ratings excluded (data from subject groups 2.A and 2.B) the

Expressor Initials	Gabor	Geometry
KA	0.782	0.574
KL	0.634	0.500
KM	0.744	0.619
KR	0.684	0.401
MK	0.644	0.512
NA	0.696	0.420
NM	0.458	0.207
TM	0.624	0.425
UY	0.653	0.206
YM	0.650	0.506
Average	0.679	0.462

Table 2. Rank correlations between model and semantic rating similarities for experiments which excluded fear stimuli and ratings.

rank correlation between Gabor model and data ranged from 0.624 (expressor TM) to 0.782 (expressor KA), with an average value of 0.679. For the geometry based control, rank correlations between model and data ranged from 0.206 (expressor UY) to 0.619 (expressor KM) with an average value of 0.462. Correlation results for all expressors are listed in Table 2. Expressor NM was considered to be an outlier and excluded from the above quoted averages and ranges. On inspection NM's expressions were difficult to interpret.

All rank correlations quoted were calculated using Spearman's method. The two sided significance of all of the deviation of all rank correlations calculated indicated a high level of significance. In all cases the correlation coefficient was greater for the Gabor model than for the model based solely on geometric displacement of feature points.

Gabor and human similarity data was analyzed using non-metric multidimensional scaling (nMDS) using the ALSCAL algorithm [13]. nMDS embeds points in a euclidean space in such a way that the distances between points preserves the rank order of the dissimilarity values between those points. "Stress" and "Rsq" respectively measure the residual misfit of the euclidean distance to the dissimilarities and the squared correlation between distances and dissimilarities. By monitering these parameters as the number of nMDS dimensions was increased, it was found that two dimensions provide an adequate embedding of the similarity data. Figs. 6 , 7 show sample nMDS solutions for human ratings similarity values and Gabor code derived similarity values. In figs. 6,7, the following abbreviations are used: NE - Neutral, HA - Happiness, SA - Sadness, SU-Surprise, AN - Anger, DI - Disgust. Fig. 5 shows sample nMDS solution in which images have been positioned at their coordinates in the euclidean space.

nMDS solutions are arbitrary up to rotation, translation and reflection of the configuration of points. In Fig. 5 the points have been rotated, translated, and reflected to show the agreement between model and data. Figs. 6,7 have not been treated in this way. The salient aspect of these plots is the relative position of the facial expression clusters.

6 Discussion

Similarity values calculated using the Gabor coding and semantic ratings showed a highly significant degree of correlation, with no parameter fitting. Non-metric multidimensional scaling uncovered a low-dimensional space in which Gabor-coded images are clustered into the known basic categories of facial expressions. Together, these findings show that this representation scheme extracts adequate image information to code the basic facial expressions. Using this input code and a perceptron classifier a facial expression recognition has been built [16]. Two sets of experiments were run, one excluding fear expressions. Model/Data agreement was higher with fear excluded. Fear is considered to be a problematic expression for Japanese expressors and subjects for reasons beyond the scope of the present article.

Interestingly, the low-dimensional spaces for ratings data and Gabor-code are similar. One axis (nearly horizontal in Fig. 5) corresponds to the degree of pleasantness (happy vs. anger and disgust) in the expression. A roughly orthogonal dimension corresponds to the level of arousal shown by the face (surprised vs. sad). This configuration was seen for all of the expressors studied (except NM, where the data is erratic). Deviations from this general arrangement visible in Figs. 6,7 are typical of other nMDS results not shown due to space limits.

The Gabor similarity showed a higher degree of correlation with the data than did a geometry-based control. Featural geometry, an explicit and precise function of facial deformation due to expression, does not capture any textural changes. Addition of more grid points could increase the performance of a geometry measure, but at the price of greatly increased computational complexity. Location of grid points is the most expensive part of a fully automatic system [7, 14]. Moreover the Gabor measure puts less stringent demands on the precision of the grid positioning, because the phase of the filter response was not used in similarity calculations. A combined Gabor+Geometry system could have still higher performance, however results of [16] indicate the improvements are minor.

Previous studies on automatic facial expression processing classify images into face expression or facial action categories. The facial images used in training or testing such systems should preferably be pictures of pure expressions posed by expert actors. A novel aspect of our work is that the system is compared with differential ratings. This

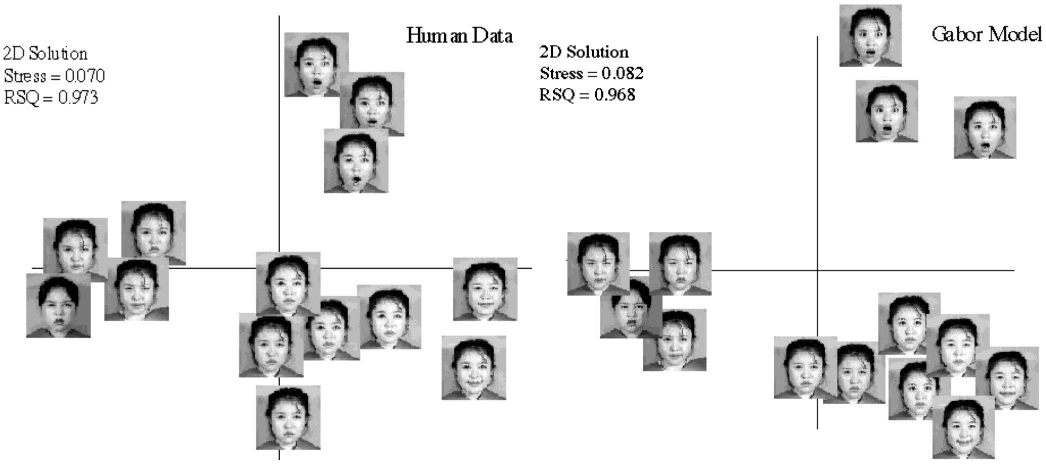

Figure 5. nMDS solutions for Gabor and semantic rating similarities (data from pilot study).

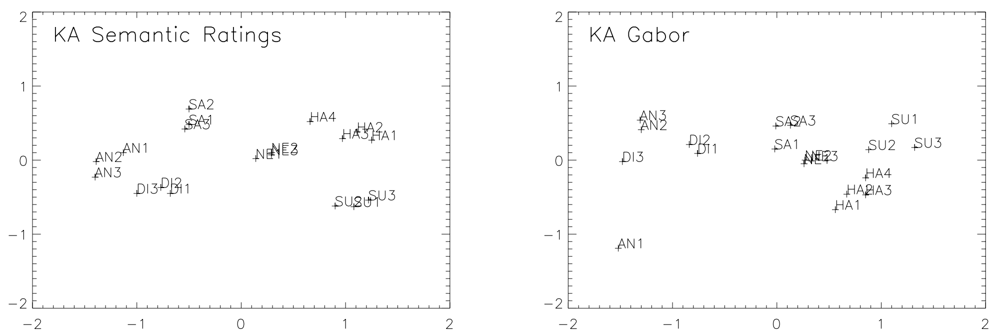

Figure 6. nMDS solutions for Gabor and semantic rating similarities (subject KA). See text for key to expression abbreviations.

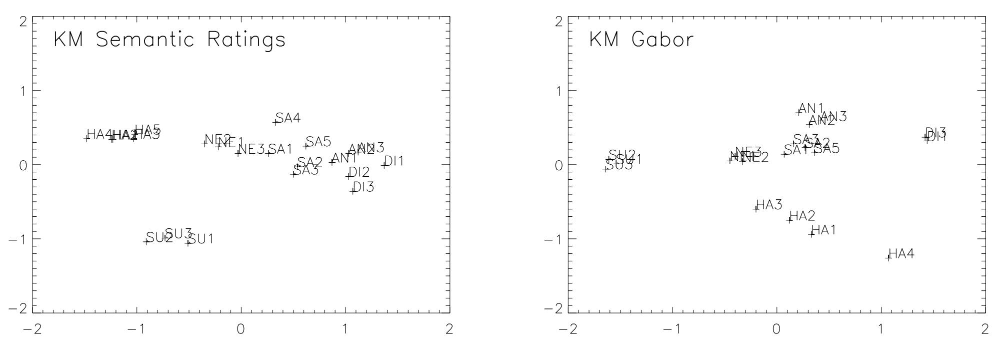

Figure 7. nMDS solution spaces for Gabor and semantic rating similarities (subject KM). See text for key to expression abbreviations.

sidesteps the requirement for pure expressions. Comparison of the system with human ratings relaxes the requirement for ground truth in labelling expression categories.

Why is there any agreement with psychology? Facial expressions are distinguished by fine changes in shape and texture of the face. From the standpoint of neurobiology, such changes are best represented using the spatially localized receptive fields typical of primary visual cortex (V1) cells. The neural systems processing facial expressions in higher vision require access to such spatially localized information. Gabor wavelet functions approximately model V1 simple cell while the amplitude of the complex Gabor transform models complex cells [3, 6, 11]. Lyons et al. [8] found that the Gabor measure predicts aspects of facial similarity perception.

Finally, it is interesting that the low-dimensional structure of the semantic ratings similarity space resembles that of the Gabor measure. Many studies in the psychological literature (beginning with Schlosberg [12]) suggest a "circumplex" arrangement of the basic facial expressions in a two-dimensional space with dimensions of pleasantness and arousal. We conjecture that high-level (even semantic level) processing of expressions may preserve organizational aspects of the low-level processing by the early visual system.

7 Acknowledgements

We thank Phi Tran for technical assistance at the beginning of the project and Reiko Kubota for help with the data collection. Pilot studies were partly supported by a grant from the Annenberg Center for Communication at University of Southern California.

References

[1] M. Bartlett, P. Viola, T. Sejnowski, J. Larsen, J. Hager, & P. Ekman. Classifying Facial Action. In: *Advances in Neural Information Processing Systems 8*, D. Touretzky et al. editors, MIT Press, Cambridge, MA, 1996.

[2] C. Padgett & G. Cottrell. Identifying emotion in static face images. In *Proceedings of the 2nd Joint Symposium on Neural Computation,* vol. 5, pp. 91-101, La Jolla, CA, 1995.

[3] J.G. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2: 1160-1169, 1985.

[4] P. Ekman & W. V. Friesen. Unmasking the Face. A guide to recognizing emotions from facial clues. Palo Alto: Consulting Psychologists Press, 1975.

[5] I. Essa and A. Pentland. Facial Expression Recognition using Visually Extracted Facial Action Parameters. In: Bichsel, M., (Ed), *Proceedings of the International Workshop on Automatic Face and Gesture Recognition* (pp. 35-40). 1995.

[6] J.P. Jones & L.A. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology* 58: 1233-1258, 1987.

[7] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, & W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture, *IEEE Transactions on Computers,* 42: 300-311, 1993.

[8] M. J. Lyons and K. Morikawa. A model based on V1 cell responses predicts human perception of facial similarity. *Investigative Ophthalmology and Visual Science,* 37: 910, 1996.

[9] C. von der Malsburg. *The Correlation Theory of Brain Function*. Internal Report 81-2, Max Planck Institute for Biophysical Chemistry, Göttingen, 1981.

[10] K. Mase. Recognition of facial expression from optical flow. *IEICE Transactions,* 74(10): 3474-3483, 1991.

[11] D. A. Pollen & S. F. Ronner. Phase relationships between adjacent simple cells in the visual cortex. *Science,* 212: 1409-1411, 1981.

[12] Schlosberg. The description of facial expressions in terms of two dimensions. *Journal of Experimental Psychology,* 44: 229-237, 1952.

[13] Y. Takane, F. W. Young, & J. de Leeuw. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42: 7-67, 1977.

[14] L. Wiskott, J. Fellous, N. Krüger, & C. von der Malsburg. Face Recognition and Gender Determination. In Bichsel, M., (Ed), *Proceedings of the International Workshop on Automatic Face and Gesture Recognition* (pp. 92-97), 1995.

[15] Y. Yacoob & L. Davis. Recognizing human facial expressions form long image sequences using optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 18: 636-642, 1996.

[16] Z. Zhang, M. Lyons, M. Schuster, & S. Akamatsu. Comparison between geometry-based and Gabor-wavelets-based facial expression recognition using multi-layer perceptron. This volume.