A Face-House Paradigm for Architectural Scene Analysis

Stephan K. Chalup*
Newcastle Robotics Lab
School of Electrical Eng. and Computer Science
The University of Newcastle
NSW 2308 Australia
stephan.chalup@newcastle.edu.au

Kenny Hong
Newcastle Robotics Lab
School of Electrical Eng. and Computer Science
The University of Newcastle
NSW 2308 Australia
kenny.hong@newcastle.edu.au

Michael J. Ostwald
School of Architecture and Built Environment
The University of Newcastle
NSW 2308 Australia
michael.ostwald@newcastle.edu.au

ABSTRACT

This interdisciplinary study proposes a method for architectural design analysis of house façades which is based on face detection and facial expression classification. The hypothesis is that abstract face expression features can occur in the architectural design of house façades and will potentially trigger emotional responses of observers. The approach used statistical learning with support vector machines for classification. In the computer experiments the system was trained using a specifically composed image data base consisting of human faces and smileys. Afterwards it was applied to a series of test images of human facial expressions and house façades. The experiments show how facial expression pattern associated with emotional states such as surprise, fear, happiness, sadness, anger, disgust, contempt or neutral could be recognised in both image data sets.

Categories and Subject Descriptors

J.5 [**Arts and Humanities**]: Architecture

General Terms

Experimentation, Human factors

Keywords

Affective computing, architecture, face recognition, image processing, support vector machines

*Corresponding author: Dr. Stephan Chalup, School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia; Phone: +61 2 492 16080; Fax: +61 2 4921 6929; Email: stephan.chalup@newcastle.edu.au

CSTST 2008 October 27-31, 2008, Cergy-Pontoise, France

1. INTRODUCTION

Various aspects of architectural design analysis have contributed to questions such as: How do we perceive aesthetics ? What determines whether a streetscape is pleasant to live in ? What factors influence our well-being when living in a particular urban neighbourhood ? Some studies proposed, for example, the involvement of harmonic ratios, others calculated the fractal dimension of façades and skylines to determine the aesthetic value of façades, street- and cityscapes [5, 7, 24, 31, 32, 43].

The present study investigates an alternative hypothesis which is inspired by results from brain research and cognitive science which show that large areas of the brain are dedicated to face processing and that communication via facial expressions involves the emotional centers of the brain [9].

Our hypothesis is that abstract face expression features can occur in the architectural design of house façades and will potentially trigger emotional responses of observers.

It is commonly known that humans have the ability to "see faces" in abstract objects which display visual features such as two dots and a line segment geometrically arranged in a configuration similar as the eyes and the nose in a face. This phenomenon has nicely been illustrated by the photographers François and Jean Robert who collected a whole book of photographs of objects which seem to display face-like structures [38].

The topic of face recognition traditionally plays an important role in cognitive science and in particular in research on object recognition and interpretation and affective computing [10, 35, 46]. A widely accepted opinion is that face recognition is a special skill and distinct from general object recognition. Farah *et al.* [11, 12, 15] proposed that faces are processed holistically (i.e. without explicit representation of parts such as eyes or mouth) and in specific areas of the human brain, the so-called fusiform face areas [13, 26]. Later studies confirmed that activation in the fusiform gyri plays a central role in the perception of faces [22, 34] and that also a number of other specific brain areas show higher activation when subjects were confronted with facial expressions than when they were shown images of neutral faces [9]. Based on recognition experiments using images of faces and houses, Farah [12] had concluded that holistic processing is more dominant for faces than for houses.

Prosopagnosia, the inability to recognise familiar faces while general object recognition is intact, was believed to be an impairment that exclusively affects a subject's abil-

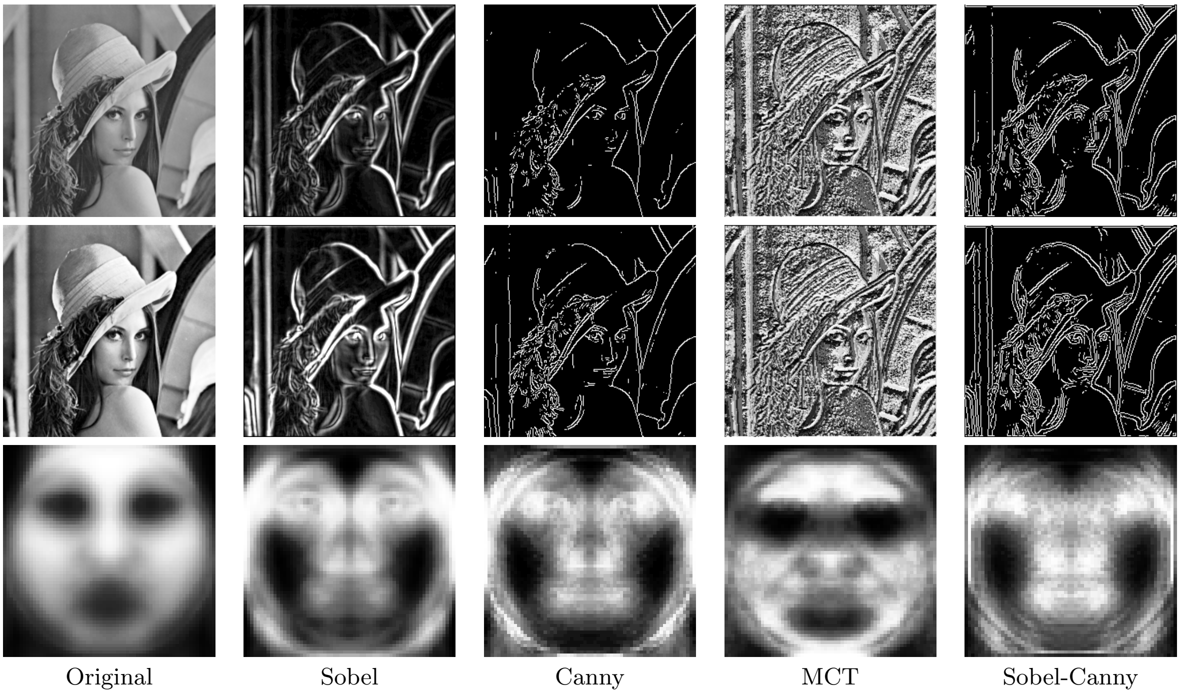

Figure 1: First row: Lenna when processed using different edge detection methods; Second row: Outcome of processing Lenna image when first histogram equalisation was applied; Third row: Average of all images of the database after histogram equalisation and application of one of the edge detection methods.

ity to recognise and distinguish familiar faces and may be caused by damage of the fusiform face areas of the brain [23]. In contrast, there is evidence which indicates that it is the expertise and familiarity with individual object categories which is associated with holistic modular processing in the fusiform gyrus and that prosopagnosia not only affects processing of faces but also of complex familiar objects [18, 19, 20, 21]. These contrasting opinions are object of an on-going discussion [14, 17, 29, 33, 37]. Although the debate is far from completed, a developmental perspective suggests the view that "the ability to recognize faces is one that is learned" [30]. The learning process starts during the first months of life where rapidly regions in the inferotemporal cortex become specialised primarily for face recognition.

The first milestone of this project was to design a face detection and emotion classification system based on statistical learning [3] and train it using a database of images of faces of human subjects. After calibrating the system's learning parameters on a separate data set of images of human face expressions we assumed the system represented a simple statistical model of how human subjects would detect and classify human faces and expressions, respectively. An evaluation of the system when applied to an image database of house façades should allow to test under which conditions the model can detect facial features and assign façade sections to human emotions.

The paper is structured as follows: In Section 2 a description of the system is given which includes modules for preprocessing, face detection and emotion classification. Results are presented and discussed in Section 3. The final section 4 is the conclusion.

2. SYSTEM AND METHOD DESCRIPTION

The aim was to design and implement a simply structured but well-performing system based on statistical learning methods. The system should not rely on domain specific techniques from human face processing such as eye and lip detection which are used in some of the current systems for biometric human face detection.

The training data should include not only faces but also abstractions of facial expressions represented by smileys. This should allow the system to learn a relatively general statistical model of facial expressions which then should not only be evaluated on human face images but also on architectural image data.

A significant part of the project addressed data selection and preparation. The final design was a modular system consisting of a preprocessing module followed by two levels of classification for face detection (one-class classifier) and emotion classification (multi-class classifier).

2.1 Face Database

The image database for training the classifiers for face detection and facial expression classification consisted of 5106 human faces and smileys that have been cropped tightly to the frontal face area to contain only the main facial features. In addition a vertically mirrored version of all images was included in the training data. Smileys were included in the training data in order to support training of the facial expression classifier on abstract features associated with different emotions. Profiles and rotated views of faces were not taken into account.

The human face images for the training set stem from several databases: Paul Ekman's Sett/Mett training cd [8], AT&T Laboratories Cambridge [1], JAFFEE: Japanese Female Facial Expression [27, 28] and FEEDTUM [44]. The smileys in the training set were taken from the online linux KDE desktop repository [2].

1828 images from Paul Ekman's Sett/Mett training cd, JAFFEE, smileys and a subset of FEEDTUM were labelled and employed for training the facial expression classifier. The remaining images were not labeled and only used to train one-class classifiers for face detection.

The images for the test set were separately selected from the Cohn- Kanade human facial expression database [25]. The underlying images printed in Figure 3 show only those subjects which gave permission according to [25].

Figure 2: Face detection and emotion classification applied to the Lenna image. The small (violet) boxes suggest "surprised", the middle (green) box which includes Lenna's concave shoulder edge suggests "disgusted", and the large (orange) boxes suggest "contempt".

2.2 Preprocessing Steps

The preprocessing module first converts all images into greyscale and then applies histogram equalization followed by edge detection.

Histogram equalization [42] compensates for effects due to changes in illumination, different camera settings, and different contrast parameters between the different images. Three options how to apply histogram equalisation were evaluated: Application to the whole image, to the content of selected boxes, or not at all. In most (but not all) cases histogram equalisation helped to detect additional relevant edges. Figure 1 shows example images with (second row) and without (first row) histogram equalisation applied. Overall we found that histogram equalisation enhanced the system's performance in particular when applied to individual boxes.

Several different edge detection methods were evaluated, including Sobel [41] edge detection, Canny [4] edge detection, the Modified Census Transform (MCT) [16], and a combination of the Sobel and Canny filters. Outcomes are shown in Figure 1. Sobel edge detection and Canny edge detection are well-known edge operators that produce grey level and binary edges, respectively. The Modified Census Transform (MCT) is a non-parametric spatial image transform [16] which produces an image that is invariant to illumination changes. The Sobel-Canny edge operator has two stages: First the Sobel stage retrieves strong grey level edges, and then the Canny stage selects the relevant edges in binary form. If the outcome of the Sobel stage included thick lines they could result in double edges after application of the Canny stage (cf. Figure 1).

2.3 Face Detection

The application of the face detection module consists of four steps plus preprocessing which can include histogram equalisation and edge detection:

1. Select a random point c within the image.
2. Select a random box size.
3. Crop the image to extract the interior of the box generated with the random point c as center. Rescale the interior of the box to a 20×20 pixel resolution.
4. Apply a one-class support vector machine classifier to decide if the box contains a face.

Central component is a one-class support vector machine (SVM) using the radial basis function (RBF) kernel. The output of the classifier is a decision value which indicates the probability that the sample belongs to the learned model class. SVMs were previously employed for face detection, e.g. by [36]. The present study employed ν-SVMs [39, 40] as implemented in the libsvm library [6]. The parameter ν was varied in the range from 0.1 to 0.5 while the γ of the RBF kernel was left at libsvm's default value of 0.0025.

For tuning the parameter ν a deterministic search scenario was applied. It tested boxes for every pixel at several different resolutions until a face was found and then it moved on to the next pixel. The smallest resolution corresponded to the SVM model's input size, which was set to 20×20 pixel. This approach generated a cloud of candidate solutions (shown as yellow clouds in this paper's figures) consisting of center points of boxes with the highest decision values output by the one-class SVM. Note that every pixel within a "face cloud" had a positive decision value (if the value was negative it meant that the pixel was not associated with a face box).

For the pilot experiments a two stage algorithm was implemented. It first detected local and then global peaks of the decision values associated with the pixels in the yellow face clouds. Local peaks were found when the surrounding pixels had a lower decision value than the center pixel. Then global peaks were determined by clusters of local peaks defined by overlapping boxes of the face(s) at these local peaks. The overlap ratios of these face boxes were compared. It was demanded that 80% of the reference box had to belong to a particular cluster. This approach allowed for small faces to appear within larger faces.

2.4 Facial Expression Classification

Affect recognition has become a large field [46]. Excellent results can be obtained through multi-modal approaches. For example, Wang & Guan [45] combined audio and visual data recognition in a system which is capable to recognize six human emotional states in human subjects with different language background with a success rate of 82%.

The purpose of the present study was to evaluate architectural image data. Therefore a purely vision based approach had to be adapted. As facial expression classifier a

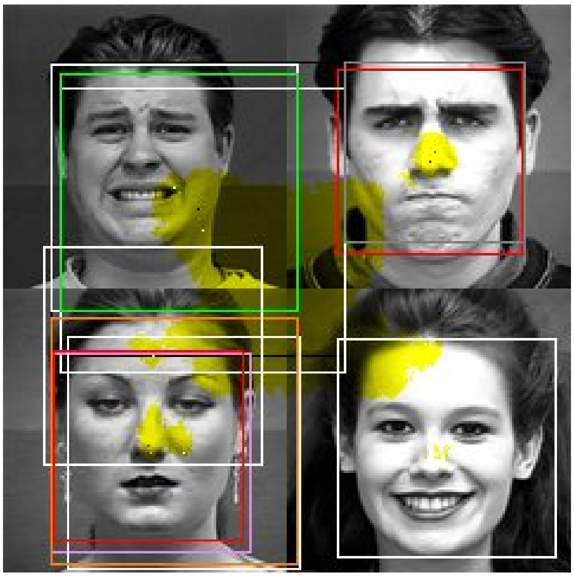

Figure 3: The trained face detection and expression classification modules were applied to test images taken from a standard database of face images [25] (©Jeffrey Cohn). The left upper expression was classified as "disgusted" or "happy", the right upper image was classified as "angry" or "neutral", the bottom right image was identified as "happy", while the left bottom picture was associated with several emotions.

multiclass ν-SVM [39] using the radial basis function (RBF) kernel was trained on the labelled data set of 1828 images (from Section 2.1). Eight classes corresponding to Ekman's facial expression classification system's (FACS) eight emotional states were distinguished [8]: surprise, fear, happiness, sadness, anger, disgust, contempt or neutral. Face expressions were colour coded via the frames of the boxes which were determined to contain a face by the face detection module in the first stage of the system. The following list describes which colours were assigned to which emotional states:

sad	=	blue
angry	=	red
surprised	=	violet
in fear	=	black
disgusted	=	green
contempt	=	orange
happy	=	white/yellow
neutral	=	grey

Examples of how the face detection and facial expression classification system was applied to test images of human faces are shown in Figures 2 and 3.

Although the face in the Lenna image (Fig. 2) was not format filling and not exactly in frontal view the system was able to detect it through box sampling. First the one-class classifier output a decision value for each box to indicate which of the candidate boxes most likely contains a face pattern. The boxes with the highest decision values were then run through the facial expression classifier. From the boxes shown in Figure 2 the boxes which most tightly fit Lenna's face were classified by the system as "surprised" (violet). Some of the other boxes with high decision values were of different size or were slightly translated. They were classified either as "disgusted" (green) or "contempt" (orange).

The test image shown in Figure 3 was composed of four face images from our test set which was not used for training. All four faces were detected as dominant face pattern by the face detection module. The facial expression classification module assigned sensible emotion classes to all images except the bottom left face which didn't express a clear emotion and resulted in several possible emotion categories including "angry" (red), "contempt" (orange), "happy" (white), and "surprised" (violet).

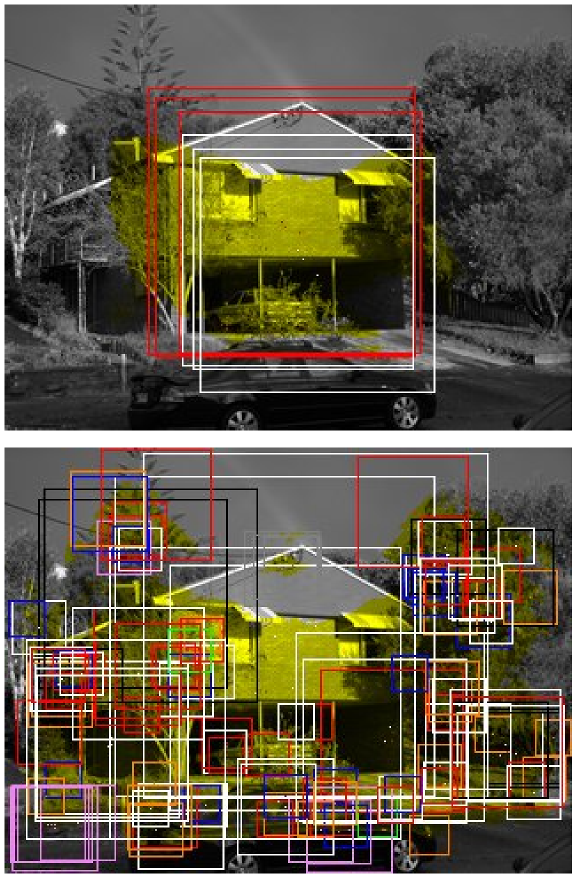

Figure 4: ***Top image:*** **The face detection module detected a face-like structure in the facade. The region captured in the upper three boxes were classified as "angry" (red) and in the lower right three boxes as "happy" (white).** ***Bottom image:*** **Appropriate preprocessing and selection of the SVM parameter ν are critical to avoid that too many faces are detected.**

3. EXPERIMENTAL RESULTS WITH ARCHITECTURAL IMAGE DATA

Using selected examples from architectural image data it was observed that sensible face detection could be performed on grayscale images without edge filtering if the contrast ratios of the house image were similar to that of the human

(a) Glebe Road

(b) Bull Street

(c) Ruit

Figure 5: Critical aspects of face detection and facial expression classification in facades: (a) An example where one dominant face is detected within the facade but the associated facial expression class depends on the size of the box; (b) Several faces are detected within the same facade; (c) The yellow face cloud has several components and face boxes with high decision values can occur in several parts of the image.

face images in the training set. For house images with different contrast ratios an approach using edge detection would work better. An example could be a white house with white windows and doors where only the frames are dark. Disadvantage of the edge filtering approach for this application appears to be lower stability which can result in many inappropriate face candidates.

A longer series of pilot tests indicated that the Sobel-Canny edge filter combined with a SVM based facial expression classifier with ν=0.12 was a good compromise for detecting faces in house façades while still performing well on the human test data. Using 10-fold cross validation the classifier with Sobel-Canny preprocessing showed 73.5% correct classification. The greyscale based classifier resulted in 76.4% on the test data with human faces but had difficulties to detect some for the human eye "obvious" faces in the tested house façades.

After the system was tuned and trained on the human face and the smiley data using the above described approach and preprocessing it was applied to a series of house images for evaluation. Figures 4 and 5 show some characteristic results of the experiments. The upper image of Figure 4 shows the façade of a house on Brooks Street in Newcastle. The face detection system when calibrated as described above indicated that the house façade contains a dominant pattern which can be classified as a face. The facial expression classifier further delivered high decision values for angry (upper left three boxes) or happy (lower right three boxes) facial expressions. However, with an inappropriate choice of preprocessing or SVM parameters either no faces were detected or too many as shown in the example in the bottom image of Figure 4.

Typically several different emotions can be detected within the same house façade. The large (grey) boxes in Figure 5 (a) indicate that the part of the image included in the boxes corresponds to a neutral face. The middle size blue boxes propose that the included face pattern belongs to the category "sad" and the small white boxes which contain a smaller fraction of the garage door as "mouth" contain the "happy" faces. In this example all ten boxes had similarly high decision values.

Evaluation of the house example with many windows in Figure 5 (b) demonstrates how different faces can be detected in the same house and that face detection and emotion classification on this type of data can be very unstable.

Figure 5 (c) shows that the yellow face cloud can have several components. In this example a sensible face pattern was detected in the central house façade and it was classified as "happy" (white) or "disgusted" (green). However, other faces, some of them with similarly high decision values, could be detected at other parts of the image. Alternative face structures could originate from texture of other façade structures but could also be caused by artifacts of the procedure which includes box cropping, resizing, antialiasing, histogram equalisation, and edge detection. The order of the individual processing steps can be changed and can have impact on the outcome of the procedure.

A possible way to obtain better stability, better precision and to avoid artifacts is to increase the image resolution of the classifier. The main experiments of this study so far employed images which were downsampled to a resolution of 20×20 pixel before input into the SVM. Figure 6 shows that an increase of the underlying resolution from 20×20 to

44×44 decreases the size of the cloud of candidate center points for face boxes from the cloud set containing all yellow and violet points to the cloud set only containing violet points.

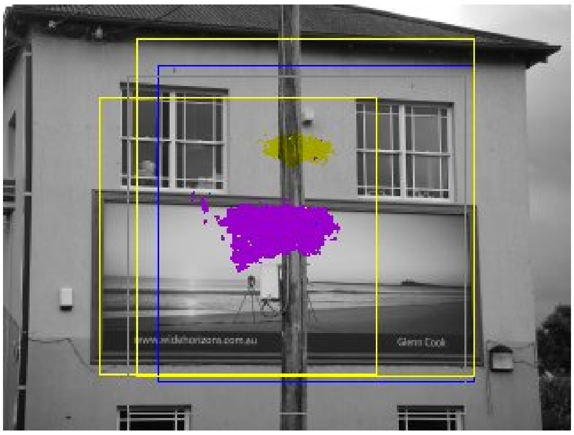

Figure 6: Results of the method depend on the image resolution. With higher resolution the cloud of candidate box center points shrinks.

4. CONCLUSION

A combined face detection and emotion classification system based on support vector classification was implemented and tested. Although the system was explicitly trained on faces and smileys it was able to detect face-expression patterns within images of selected houses. Most "faces" detected in houses were very abstract and allowed the assignment of several different emotion categories depending on the choice of the center point and the viewing angle. The pilot experiments of the present study indicated that for selected houses a dominant emotion category is identifiable. The outcome of this study supports the hypothesis that face recognition is critical for how humans perceive and interpret their environment. This includes how humans perceive the aesthetics and architecture of house façades of the buildings they live in and are surrounded by in their day-to-day lives. The presented pilot results are part of an ongoing interdisciplinary study between architecture and computer science.

5. ACKNOWLEDGEMENTS

This project was supported by ARC discovery grant DP0770106 "Shaping social and cultural spaces: the application of computer visualisation and machine learning techniques to the design of architectural and urban spaces".

6. REFERENCES

[1] AT&T Laboratories Cambridge face database. http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[2] Emoticons database at kde-look. http://www.kde-look.org/.

[3] C. M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 2006.

[4] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–714, 2001.

[5] S. K. Chalup, N. Henderson, M. J. Ostwald, and L. Wiklendt. A method for cityscape analysis by determining the fractal dimension of its skyline. In *Anzasca, November 2008, Newcastle, Australia*, 2008. forthcoming.

[6] C.-C. Chang and C.-J. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at www.csie.ntu.edu.tw/∼cjlin/libsvm.

[7] J. C. Cooper. *The potential of chaos and fractal analysis in urban design*. Joint Centre for Urban Design, Oxford Brookes University, Oxford, 2000.

[8] P. Ekman, W. V. Friesen, and J. C. Hager. *Facial Action Coding System, The Manual*. A Human Face, 666 Malibu Drive, Salt Lake City UT 84107, 2002.

[9] A. D. Engell and J. V. Haxby. Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia*, 45(14):323–341, 2007.

[10] M. W. Eysenck and M. T. Keane. *Cognitive Psychology: A Student's Handbook*. Taylor & Francis, 2005.

[11] M. J. Farah. *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision*. The MIT Press, Cambridge, MA, 1990.

[12] M. J. Farah. Neuropsychological inference with an interactive brain: A critique of the 'locality assumption'. *Behavioral and Brain Sciences*, 17:43–61, 1994.

[13] M. J. Farah and G. K. Aguirre. Imaging visual recognition: PET and fMRI studies of the functional anatomy of human visual recognition. *Trends in Cognitive Sciences*, 3(5):179–186, May 1999.

[14] M. J. Farah, C. Rabinowitz, G. E. Quinn, and G. T. Liu. Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology*, 17:117–124, 2000.

[15] M. J. Farah, K. D. Wilson, M. Drain, and J. N. Tanaka. What is "special" about face percpetion? *Psychological Review*, 105:482–498, 1998.

[16] B. Fröba and A. Ernst. Face detection with the modified census transform. In *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)*, 2004.

[17] I. Gauthier and C. Bukach. Should we reject the expertise hypothesis? *Cognition*, 103(2):322–330, 2007.

[18] I. Gauthier, T. Curran, K. M. Curby, and D. Collins. Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience*, (6):428–432, 2003.

[19] I. Gauthier, P. Skudlarski, J. C. Gore, and A. W. Anderson. Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, 3:191–197, 2000.

[20] I. Gauthier and M. J. Tarr. Becoming a "greeble" expert: Exploring face recognition mechanisms. *Vision Research*, 37:1673–1682, 1997.

[21] I. Gauthier, M. J. Tarr, A. W. Anderson, P. Skudlarski, and J. C. Gore. Activation of the middle fusiform face area increases with expertise in

recognizing novel objects. *Nature Neuroscience*, 2:568–580, 1999.

[22] N. George, J. Driver, and R. J. Dolan. Seen gaze-direction modulates fusiform activity and its coupling with other brain areas during face processing. *NeuroImage*, 13(6):1102–1112, 2001.

[23] N. Hadjikhani and B. de Gelder. Neural basis of prosopagnosia: An fMRI study. *Human Brain Mapping*, 16:176–182, 2002.

[24] C. M. Hagerhall, T. Purcell, and R. P. Taylor. Fractal dimension of landscape silhouette as a predictor of landscape preference. *The Journal of Environmental Psychology*, 24:247–255, 2004.

[25] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (FG'00), Grenoble, France*, pages 4–53, 2000.

[26] N. Kanwisher, J. McDermott, and M. M. Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17:4302–4311, 1997.

[27] M. J. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. In *Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society*, pages 200–205, 1998.

[28] M. J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, 1999.

[29] E. McKone and R. A. Robbins. The evidence rejects the expertise hypothesis: Reply to Gauthier & Bukach. *Cognition*, 103(2):331–336, 2007.

[30] C. A. Nelson. The development and neural bases of face recognition. *Infant and Child Development*, 10:3–18, 2001.

[31] M. J. Ostwald, J. Vaughan, and S. Chalup. A computational analysis of fractal dimensions in the architecture of Eileen Gray. In *ACADIA 2008, Silicon + Skin: Biological Processes and Computation, October 16 - 19, 2008*, 2008.

[32] M. J. Ostwald, J. Vaughan, and C. Tucker. Characteristic visual complexity: Fractal dimensions in the architecture of Frank Lloyd Wright and Le Corbusier. In K. Williams, editor, *Nexus: Architecture and Mathematics*, pages 217–232. Turin: K. W. Books and Birkhäuser, 2008.

[33] T. J. Palmeri and I. Gauthier. Visual object understanding. *Nature Reviews Neuroscience*, 5:291–303, 2004.

[34] K. A. Pelphrey, J. D. Singerman, T. Allison, and G. McCarthy. Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia*, 41(2):156–170, 2003.

[35] R. W. Picard. *Affective Computing*. Cambridge: The MIT Press, 1997.

[36] T. Poggio, B. Heisele, and P. Ho. Face recognition with support vector machines: Global versus component-based approach. In *Proceedings of the Eighth IEEE International Conference on Computer Vision*, volume 2, pages 688–694, 2001.

[37] R. A. Robbins and E. McKone. No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition*, 103(1):34–79, 2007.

[38] F. Robert and J. Robert. *Faces*. Chronicle Books, San Franciso, 2000.

[39] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett. New support vector algorithms. *Neural Computation*, 12:1207–1245, 2000.

[40] B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.

[41] I. E. Sobel. *Camera Models and Machine Perception*. Ph.d. dissertation, Stanford University, Paolo Alto California, 1970.

[42] K. K. Sung. *Learning and Example Seletion for Object and Pattern Detection*. Ph.d. dissertation, Massachusetts Institute of Technology, 1996.

[43] R. P. Taylor. Reduction of physiological stress using fractal art and architecture. *Leonardo*, 39(3):25–251, 2006.

[44] F. Wallhoff. Facial expressions and emotion database. http://www.mmk.ei.tum.de/ waf/fgnet/feedtum.html, 2006. Technische Universität München.

[45] Y. Wang and L. Guan. Recognizing human emotional state from audiovisual signals. *IEEE Transactions on Multimedia*, 10(4):659–668, June 2008.

[46] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008. accepted.