massachusetts institute of technology — artificial intelligence laboratory

Detecting Faces in Impoverished Images

Antonio Torralba and Pawan Sinha

AI Memo 2001-028 November 2001
CBCL Memo 208

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

The ability to detect faces in images is of critical ecological significance. It is a pre-requisite for other important face perception tasks such as person identification, gender classification and affect analysis. Here we address the question of how the visual system classifies images into face and non-face patterns. We focus on face detection in impoverished images, which allow us to explore information thresholds required for different levels of performance. Our experimental results provide lower bounds on image resolution needed for reliable discrimination between face and non-face patterns and help characterize the nature of facial representations used by the visual system under degraded viewing conditions. Specifically, they enable an evaluation of the contribution of luminance contrast, image orientation and local context on face-detection performance.

Research reported in this paper was supported in part by funds from the Defense Advanced Research Projects Agency and a Sloan fellowship for neuroscience to PS.

1. INTRODUCTION

One of the most salient aspects of the human visual system is its ability to robustly interpret images under conditions that drastically reduce the effective image resolution. Probing the limits of this ability can yield valuable insights regarding the nature of the representations that the visual system uses for specific recognition tasks.

In this paper, we focus on the task of face-detection under impoverished viewing conditions – determining whether an image pattern is a human face or not. Classifying an image fragment as a face is a necessary first step for many other facial analyses including identification, gender classification and affect recognition. Our emphasis on impoverished viewing conditions is motivated by three factors. First, normal viewing conditions are rarely optimal. Viewing distances may be large, the optics of the eyes may have refractive errors and the transparency of the atmosphere may be compromised by haze or smoke. Second, it may be easier to determine critical attributes necessary for face detection by reducing the amount of information available in an image. Third, experiments with impoverished images also implicitly allow us to characterize the performance of people with low-vision. Such information is valuable for developing rehabilitation programs and devices.

A more pragmatic motivation for undertaking these studies derives from the domain of computer vision. The human visual system often serves as the de-facto standard for evaluating machine vision approaches. This is particularly true in the domain of face recognition where the versatility and robustness of human recognition mechanisms implicitly define the performance goals that artificial systems seek to match and eventually exceed. Clearly, in order to be able to use the human visual system as a useful standard to strive towards, we need to first have a comprehensive characterization of its capabilities. Considering the ecological significance of detecting faces at a distance, we can expect evolution to have endowed the primate brain with powerful strategies for accomplishing this task. Knowing the limits of performance of these recognition strategies under different conditions and with different cues can allow us to evaluate the potential of different proposed computer vision approaches and also how well their performance approaches the standard. It is important to stress that the limits of human performance do not necessarily define upper bounds on what is achievable. Specialized person detection systems (say those based on novel sensors, such as IR cameras) may well exceed human performance. However, in many real-world scenarios using conventional sensors, matching human performance remains an elusive goal. We hope that our experiments can not only give us a better sense of what this goal is, but also what computational strategies we could employ to move towards it and, eventually, past it.

Surprisingly, there has been very little experimental work so far on face-detection. Most of the research attention has been directed to face-identification. Pioneering work on face identification with low-resolution imagery was done by Harmon and Julesz [1973] and Morrone et al [1983]. Working with block averaged images of familiar faces (of the kind shown in figure 1), they found high recognition accuracies (approximately 95%) even with images containing just 16x16 blocks. More recently, Bachmann [1991] and Costen et al. [1994] have presented data that shows the dependence of face identification performance on facial images with systematically varied resolution.

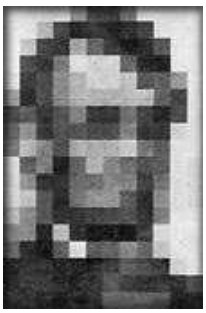

Figure 1. Images such as the one shown here have been used by several researchers to assess the limits of human face identification processes.

In all of these studies, the images presented were exclusively of faces. The experiments were designed to study within-class discrimination ('whose face is it?') rather than face classification per se ('is this a face?'). Consequently, no systematic data exist about the dependence of face-detection performance on key image attributes such as resolution, contrast polarity and orientation. We have conducted a series of experiments to address these issues with the goal of characterizing the nature of facial representations used by the human visual system.

2. FACE DETECTION EXPERIMENTS

The specific questions we have investigated in this study are:

Experiment 1: How does face detection accuracy change as a function of available image resolution?
Experiment 2: Does the inclusion of local context around faces improve face detection performance?
Experiment 3: How, if at all, do contrast negation and image orientation changes affect face detection?

To be able to conduct these experiments, we have to confront an interesting challenge - what patterns should we use as non-faces? Selecting random fragments from non-face images is not a well-controlled approach. The face/non-face discrimination can be rendered unnaturally easy for certain choices of non-face images (for instance, imagine drawing non-face patterns from a sky image). We need a more principled approach to generating non-face patterns.

In very general terms, we would like to be able to draw our non-face patterns from the same general area in a high-dimensional object space where the face patterns are clustered. Morphing between face and non-face patterns is not a satisfactory strategy since all the intermediate morphs do have a contribution from a genuine face pattern and cannot, therefore, be considered true non-faces. An alternative strategy lies in using computational classification systems that operate by implicitly encoding clusters in multidimensional spaces [Yang & Huang, 1994; Sung and Poggio, 1994; Rowley et al, 1995]. Non-face patterns on which such systems make mistakes can then serve as the distractors for our psychophysical tasks. This approach, though not entirely devoid of shortcomings, is the one we have used in our work. The key caveat to keep in mind here is that the multidimensional cluster implicitly used by these computational systems may be different from the cluster encoded by the human visual system. However, based on the high-level of classification accuracy that at least some of these systems exhibit, it is reasonable to assume that there is a significant amount of congruence between the clusters identified by them and human observers.

2.1. Experiment 1: Face detection at low-resolution

What is the minimum resolution needed by human observers to reliably distinguish between face and non-face patterns? More generally, how does the accuracy of face classification by human observers change as a function of available image resolution? These are the questions our first experiment is designed to answer. The study of images degraded due to blur provides a measure of the amount of information that is required for solving the detection task.

2.1.1. Methods:

Subjects were presented with randomly interleaved face and non-face patterns and, in a 'yes-no' paradigm, were asked to classify them as such. The stimuli were grouped in blocks, each having the same set of patterns, but at different resolutions. The presentation order of the blocks proceeded from the lowest resolution to the highest. Ten subjects participated in the experiment. They were drawn from undergraduate and graduate student populations at MIT and had normal or corrected to normal acuity. Presentations were self-timed and the images stayed up until the subject had responded by pressing one of two keys (one for 'face' and the other for 'non-face'). Stimuli were presented on a 19" Sony Trinitron monitor connected to a PIII 750 MHz PC running Windows 2000.

Our stimulus set comprised 200 monochrome patterns. Of these, 100 were faces of both genders under different lighting conditions (set 1), 75 were non-face patterns (set 2) derived from a well-known face-detection program (developed at the Carnegie Mellon University by Rowley et al [1995]) and the remaining

25 were patterns selected from natural images that have similar power-spectra as the face patterns (set 3). The patterns included in set 2 were false alarms (FAs) of Rowley et al's computational system, corresponding to the most conservative acceptance criterion yielding 95% hit rate. Sample non-face images used in our experiments are shown in figure 2. All of the face images were frontal and showed the face from the middle of the forehead to just below the mouth. Reduction in resolution was accomplished via convolution with Gaussians of different sizes (with standard deviations set to yield 2, 3, 4, and 6 cycles per face; these correspond to 1.3, 2, 2.5 and 3.9 cycles within the eye-to-eye distance ('ete'). All spatial resolutions henceforth are reported in terms of number of cycles between the two eyes).

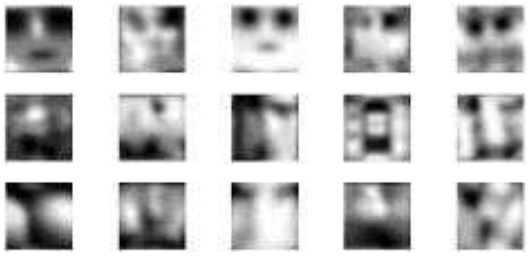

Figure 2. A few of the non-face patterns used in our experiments. The patterns comprise false alarms of a computational face-detection system and images with similar spectra as face images.

From the pooled responses of all subjects at each blur level, we computed the mean hit-rate for the true face stimuli and false alarm rates for each set of distractor patterns. These data indicated how subjects' face-classification performance changed as a function of image resolution. Also, for a given level of performance, we were able to determine the minimum image resolution required.

2.1.2. Results:

Figure 3 shows data averaged across 10 subjects. Subjects achieved a high hit rate (96%) and a low false-alarm rate (6% with Rowley et al's FPs and 0% with the other distractors) with images having only 3.9 cycles between the eyes. Performance remained robust (90% hit-rate and 19% false-alarm rate with the Rowley et al's FA distractor set) at even higher degrees of blur (2 cycles/ete). In proceeding from 2 to 1.3 cycles/ete, the hit-rate fell appreciably, but subjects were still able to reliably distinguish between faces and non-faces.

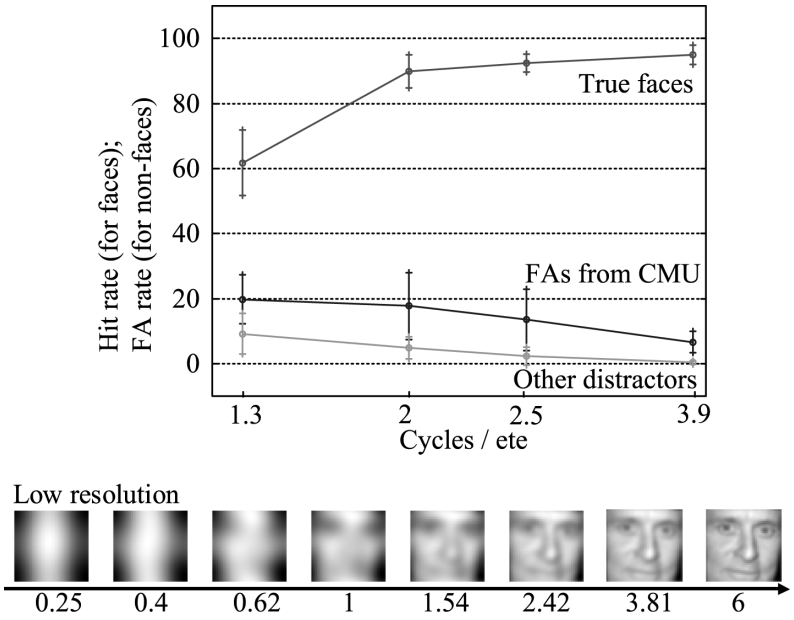

Figure 3. Results from experiment 1.The resolution units are the number of cycles eye to eye (ete).

To the best of our knowledge, this is the first systematic study of face-detection across multiple resolutions. The data provide lower-bounds on image-resolution sufficient for reliable discrimination between faces and non-faces. They indicate that the internal facial representations encode, and can be matched against, facial image fragments containing merely 2 cycles between the two eyes (cycles/eye to eye). The data also show that even under highly degraded conditions, humans are correctly able to reject most non-face patterns that the artificial systems confuse for faces. To further underscore the differences in capabilities of current computational face detection systems and the HVS, it is instructive to consider the typical image resolution needed by a few of the proposed machine-based systems: 19x19 pixels for Sung and Poggio [1994]; 20x20 for Rowley et al [1995]; 24x24 for Viola and Jones [2001] and 58x58 for Heisle et al. [2001]). Thus, computational systems not only require a much larger amount of facial detail for detecting faces in real scenes, but also yield false alarms that are correctly rejected by human observers even at resolutions much lower than what they were originally detected at.

Impressive as this performance of the HVS is, it may be an underestimate of observers' capabilities. It is possible that the inclusion of context can improve performance further. In other words, in experiment 1, subjects made the face vs. non-face discrimination on the basis of the internal structure of faces. It has traditionally been assumed that this is the pattern that defines a face. However, it is not known whether the human visual system can use the local context around the internal features to improve its discrimination and to better tolerate image resolution reductions. Experiment 2 addresses this issue.

2.2. Experiment 2: The role of local context in face-detection

The prototypical configuration of the eyes, nose and mouth (the 'internal features') intuitively seems to be the most diagnostic cue for distinguishing between faces and non-faces. Indeed, machine based face detection systems typically rely exclusively on internal facial structure [Sung & Poggio, 1994; Rowley et al., 1995; Leung et al. 1995]. External facial attributes such as hair, facial bounding contours and jaw-line are believed to be too variable across individuals for inclusion in a stable face representation. These attributes constitute the local context of internal facial features. To assess the contribution of local context to face-detection, we repeated experiment 1 with image fragments expanded to thrice their sizes in each dimension (see figure 4). The experimental paradigm was the same as for experiment 1. Subject pools for experiments 1 and 2 were mutually exclusive.

Figure 4. Faces (left set) and non-faces (right set) with local context.

2.2.1. Results:

We tested 10 subjects on the 'expanded' version of images used in experiment 1. Figure 5 shows the results. Performance improved significantly following this change. Faces could be reliably distinguished from non-faces even with just 4 cycles across the entire image (which translates to 0.87 cycles/ete). At this resolution, the internal facial features become rather indistinct and, as the results from experiment 1 suggest, they lose their effectiveness as good predictors of whether a pattern is a face or not. It is also important to note that the contextual structure across different stimuli used in this experiment is very different. Faces were photographed against very different backgrounds and no effort was made to normalize the appearance of the context. Given that there is not enough consistent information within the face or outside of it for reliable classification, the likely explanation for the human visual system's impressive performance is that bounding contour information is incorporated in facial representations used for detection. As figure 5 shows, for comparable levels of performance, the use of bounding contours nearly halves the resolution lower-bounds needed for distinguishing faces from non-faces relative to the internal features only condition. Thus, the inclusion of bounding contours allows for tolerance to greater refractive errors in the eyes and/or longer viewing distances. This result also provides a useful hint for the

design of artificial face detection systems. By augmenting their facial representation to include bounding contours, computational systems can be expected to improve their performance markedly.

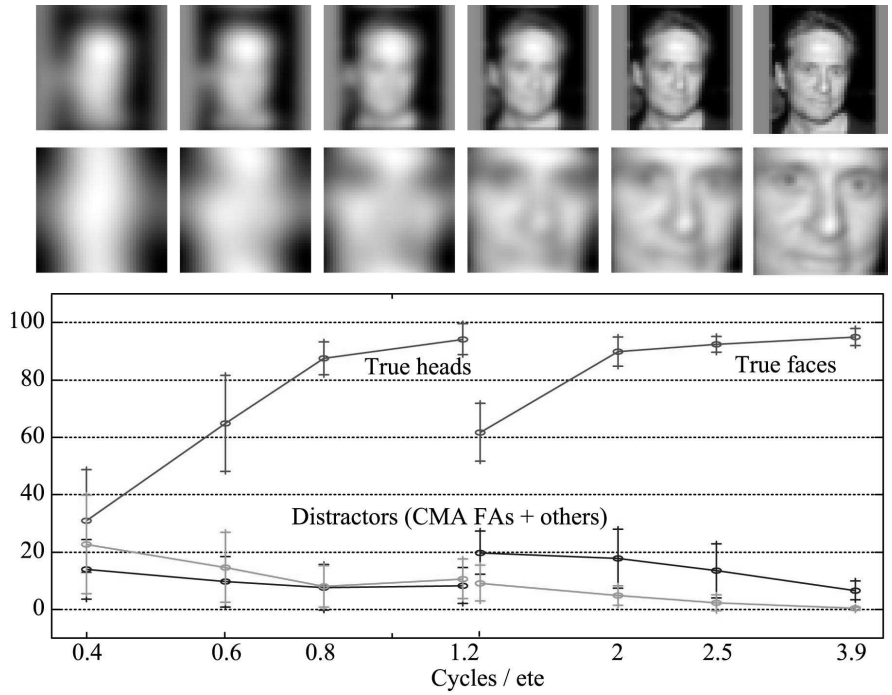

Figure 5. Results from experiment 2.

In summary, experiments 1 and 2 allow us to systematically characterize human face detection performance at very low resolutions and demonstrate a remarkable tolerance of the face-detection processes to severe reductions in resolution. Besides suggesting that both internal and external features contribute to facial encoding, the results also allow us to demarcate zones along the resolution dimension within which the two kinds of features are most effective.

Having characterized face-detection performance as a function of resolution, we next explore the roles of two other key image attributes – contrast polarity and orientation.

2.3. Experiment 3: Role of contrast polarity and face orientation in face detection

In studies of face identification, it has been found that contrast negation and vertical inversion have significant detrimental effects on performance [Galper, 1970; Bruce & Langton, 1994]. These findings have allowed researchers to make important inferences regarding the nature of facial information used for making identity judgments. However, it is unknown what role these factors play in the face-detection task. A priori, it is not clear whether these transformations should have any detrimental effects at all. For instance, it may well be the case that though it is difficult to identify people in photographic negatives or in mis-oriented images, the ability to say whether a face is present may be unaffected. Experiment 3 is designed to test this issue. The basic experimental design follows from experiments 1 and 2. However the stimulus set of experiment 3 was augmented to include additional stimuli showing the faces and non-faces contrast negated, inverted and both (figure 6 shows a few stimuli). We expected contrast negation to have little or no effect on face detection performance since this operation preserves the basic geometry of the face. As for vertical inversion, past research [Tong et al, 2000] has presented some preliminary data suggesting that this transformation has negligible impact on face-detection performance. The results we describe below show that our expectations regarding both of these transformations need to be revised.

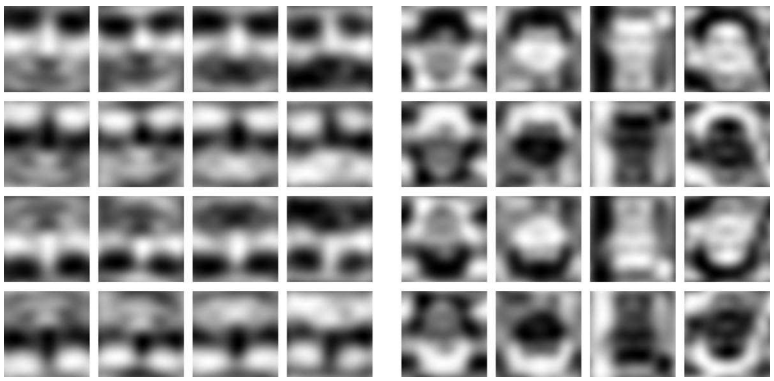

Figure 6. Stimuli that have been contrast negated and/or vertically inverted (3 cycles / image).

2.3.1. Results

Figure 7 shows results from the 'internal features only' and 'faces with local context' conditions respectively. Both contrast negation and vertical inversion influenced detection performance. Interestingly, we got very different patterns of results in the two conditions. While for internal faces, contrast negation had a much greater detrimental effect than vertical inversion, the two had about equal effects when local context was included; contrast negation overall had a smaller influence on performance with local context than without it. It appears that is that the existence of facial bounding contours renders subjects' performance more robust against contrast negation.

It is interesting that contrast negation has a strong detrimental effect on detection performance with internal features given that this transformation leaves the geometric information unchanged. The reason may lie in the statistics of the stimuli that the visual system encounters. Since, in the real world, faces have strong photometric structure (for instance the regions of the eye is systematically darker than forehead, nose and cheeks [Thoresz & Sinha, 2001; Sadr et al., 2001]), those regularities are diagnostic of face patterns and should play a mayor role in the internal representation of a face pattern. Contrast reversal of face patterns destroys the diagnostic information that allows detecting low-resolution faces. In order to be able to classify a contrast-reversed face as a face, it is necessary to increase the resolution so that the individual face features can be identified. However, in the case of facial bounding contours, the inputs mandate insensitivity to contrast polarity since faces can appear against light or dark backgrounds.

Also surprisingly, prior knowledge of the transformation did not influence the results. Half the subjects were told beforehand that the faces may appear contrast negated and/or vertically inverted. Data from the two populations were not significantly different. It appears that cognitive knowledge about possible transformations is of limited use for at least this pattern classification task.

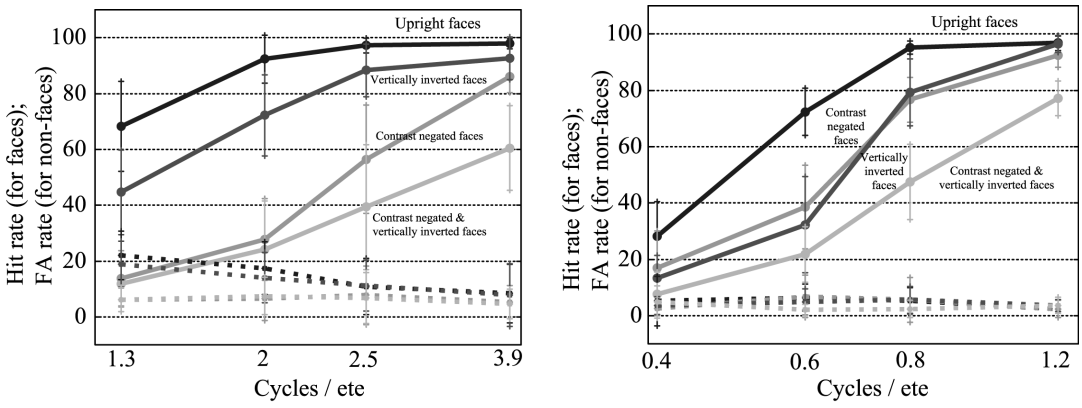

Figure 7. Results from contrast negated and inverted stimuli without (left panel) and with (right panel) context. Axis and line-labels are the same for the two panels (same color code for the graphs).

To better characterize the influence of face orientation on detection performance, we also conducted experiments with graded changes in orientation. Data in figure 8 show hit and false alarm rates for internal faces and heads as functions of image orientation (under normal and contrast reversed conditions) averaged across 20 observers. Specifically, we were interested in determining whether misorientations along certain axes were particularly disruptive for performance. Vertical bilateral symmetry is often considered an important defining attribute of faces [Reisfeld and Yeshurun, 1992; Thornhill and Gangested, 1993; Sun et al, 1998]. We expected, therefore, that detection performance would be disrupted disproportionately for orientations that destroyed the vertical bilateral symmetry. However, we found no statistically significant evidence in support of this hypothesis. The data show a graded decrease in performance as the orientation rotates away from the vertical. It is possible, however, that bilateral symmetry per se, without the requirement of the axis of symmetry being vertical, may be a determinant of face detection performance. We are undertaking experiments that explicitly manipulate facial symmetry to determine its role in face detection.

Figure 9 summarizes the results from this experiment by showing information requirements for achieving 80% correct performance (considering both hits and correct rejections) using inner only or inner and external features as a function of orientation and contrast polarity. When using both inner and external features (lower blue and green curves), contrast inversion does not significantly change the resolution required to attain 80% of performance. However, when using only inner facial features, contrast inversion has a large effect and increases the resolution-demands by more than 200% in order to have enough information to be able to compensate for the anomalous photometric distribution.

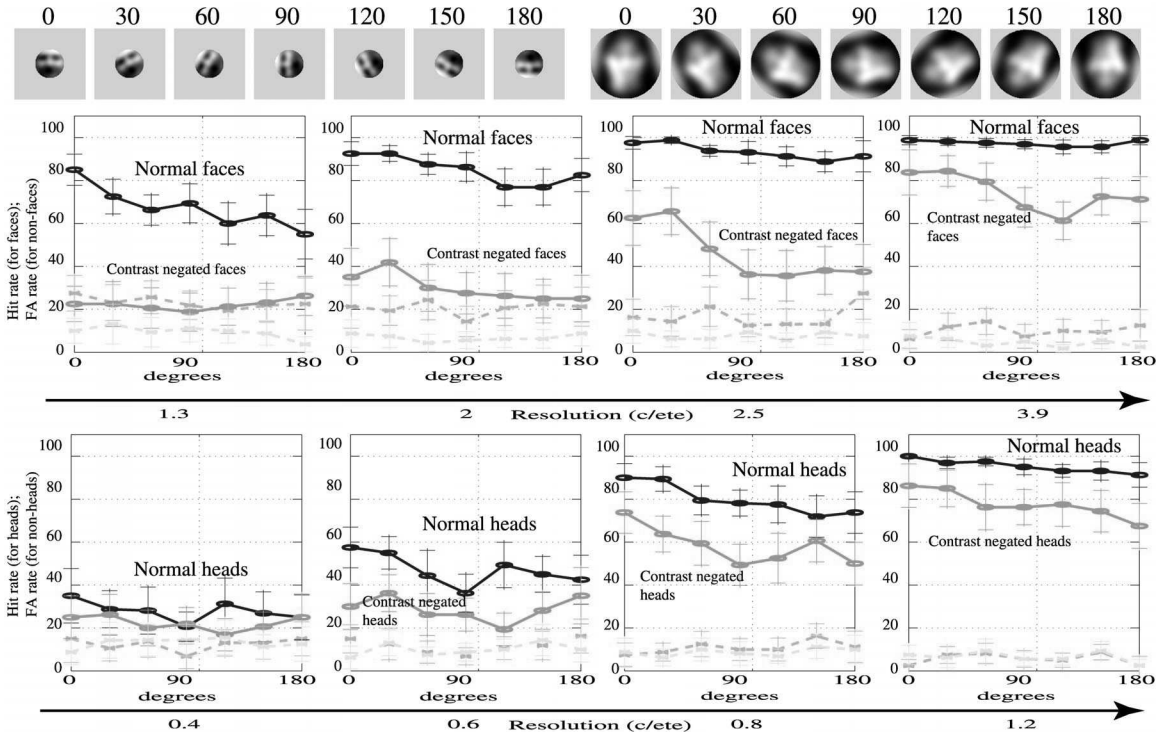

Figure 8. Influence of orientation on face detection performance. From left to right, top row shows data corresponding to 1.3, 2, 2.5 and 3.9 cycles/ete conditions for the inner facial features alone, and bottom row shows data corresponding to 0.4, 0.6, 0.8 and 1.2 cycles/ete for inner and external facial features. The blue and yellow curves correspond to the false alarm rates with distractor patterns.

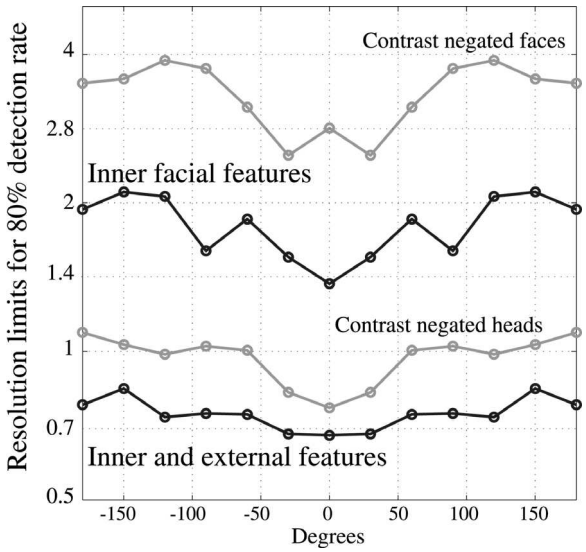

Figure 9. Information requirements for 80% of performance (considering both correct detected and correct rejected patterns) using inner only or inner and external features as a function of orientation and contrast polarity. See text for details.

GENERAL DISCUSSION:

We have conducted three experiments with the goal of characterizing the nature of representations used by human observers for classifying patterns as faces or non-faces. Our experimental results allow us to derive the following inferences:

1. The lower bounds on image resolution needed for a particular level of face-detection performance: Faces can be reliably distinguished from non-faces even at just 1.3 cycles eye-to-eye using only the internal facial information. We can also demarcate zones on the resolution axis where specific facial attributes (internal features, bounding contours) suffice for achieving a given level of detection performance.
2. The role of local context in face-detection: The inclusion of facial bounding contour substantially improves face detection performance, indicating that the internal facial representations encode this information.
3. The role of luminance contrast polarity: Contrast polarity is encoded in the representation since polarity reversals have significant detrimental effects on detection performance, particularly with inner features. The visual system is more tolerant to contrast negation in the presence of bounding contours perhaps by encoding these contours in a contrast invariant manner.
4. The role of image orientation: Changes in image orientation away from the upright decrease face detection performance. Given the largely monotonic decrease in going from upright to vertically inverted faces, our data do not support the idea that vertical bilateral symmetry per se may be a significant determinant of face detection performance.

Besides helping us characterize the nature of facial representations, these data may also allow us to address some important issues regarding the neural bases of face-detection. By employing our stimulus set in an imaging or single unit recording setting, we can obtain relative levels of neural activation for different image transformations (such as vertical inversion or contrast polarity reversal). Co-modulation of neural activity with behaviorally observed data as a function of the different transformations would allow us to infer which cortical sites, besides those already identified [Kanwisher et al, 1997], may be involved in the task of face-detection.

Several open questions remain. Although so far we have focused on face detection using full facial images, under some circumstances, classification may need to rely on fragmentary information. Partially occluded

faces constitute one such situation. The importance of fragmentary information is also highlighted by configurally deviant facial images such as those that Picasso often included in his paintings (figure 10). In these situations, the severely distorted facial geometry likely induces a greater reliance on the individual parts rather than their mutual relationships. It will be interesting to investigate how well observers are able to recognize partial fragments of faces as a function of image resolution.

Figure 10. A face by Picasso. Individual features are more diagnostic than their overall (unnatural) configuration.

The lower-bounds on image-resolution for individual features can be translated into effective whole-face resolutions so that one may directly compare this data with that from experiments 1 and 2. This comparison will allow us to demarcate zones along the resolution axis where the information used by the visual system is exclusively overall configuration based and those where it may be both configuration and parts based. Such a distinction would be invaluable for future studies in the developmental domain (children's use of configural or featural information) [Mondloch et al. 1999] and in neurophysiology (is the response of an area/cell cue invariant or driven primarily by configural or featural information?).

Also interesting would be an assessment of face-detection performance as a function of eccentricity. Based on the available data regarding how acuity changes away from the fovea, we can predict how face-detection performance should decline with increasing eccentricity. It would be interesting to determine whether actual data do indeed match predicted levels of performance or whether the adaptive significance of face detection has led to heightened sensitivity to facial patterns in the periphery.

The task of face-detection, besides being interesting in its own right, also serves as a launching pad for many other important investigations. First, how do the resolution requirements for face-detection compare to those for other face-perception tasks such as face identification, emotion recognition and gender classification? We have begun exploring this question in a series of experiments and the results will be described in a forthcoming publication [Torralba and Sinha, in preparation]. Second, in images even more highly degraded than the ones we have considered here, how does the visual system perform the task of person detection? Figure 11 illustrates the problem. The people in the image on the left are so small that detecting them by their facial structure or even body shape is not a tenable strategy. In such circumstances, contextual cues appear to be more important than the highly impoverished intrinsic object cues. We are addressing this problem by psychophysically estimating the contribution of scene-context to person detection performance and developing a computational model of contextual influences on object detection [Torralba and Sinha, 2001]. The model is yielding promising results such as the one shown in the right panel of figure 11. It is able to localize people in the image based on contextual rather than intrinsic object cues.

Figure 11. Person detection in large scenes where intrinsic information about faces and bodies may be highly impoverished. Left panel: a sample scene. Right panel: Results from a computational model for incorporating context in object detection tasks. Selection of image regions with high priors about people presence.

Our data are beginning to allow us to benchmark face-detection performance of the human visual system by systematically characterizing the consequences of key image transformations. They already point towards important clues regarding the nature of internal facial representations and serve as guides in our attempts at creating better computational models for face and person-detection.

ACKNOWLEDGEMENTS:

The authors would like to thank Sharon Gilad, Janice Chen, Dr. Aude Oliva, Profs. Tomaso Poggio and Takeo Kanade for helpful discussions regarding this work.

REFERENCES

Bachmann, T. (1991). Identification of spatially quantized tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, 3, 85-103.

Bruce, V. (1994) Stability from variation: the case of face recognition. *Quarterly Journal of Experimental Psychology*, 47A, 5-28.

Bruce, V. & Langton, S. (1994). The use of pigmentation and shading information in recognizing the sex and identities of faces. *Perception,* 23, 803-822.

Cabeza, R., Bruce, V., Kato, T. & Oda, M. (1998). The prototype effect in face recognition: Extension and limits. *Memory and Cognition.*

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. *Perception*, 23, 129-146.

DeAngelis, G. C., Ohzawa, I. & Freeman, R. D. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. General characteristics and postnatal development. *J. Neurophysiol.* 69: 1091-1117.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science*, 19, 207-208.

Harmon, L. D. & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional noise. *Science*, 180, 1194-1197.

Haritaoglu, I., Harwood, D. & Davis, L. S. (1998). Who? When? Where? What? A real time system for detecting and tracking people. *Proceedings of the Third International Conference on Face and Gesture Recognition*, Nara, Japan.

Heisle, B., T. Serre, S. Mukherjee and T. Poggio. (2001) Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images. In: *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, IEEE Computer Society Press, Jauai, Hawaii, December 8-14, 2001, (to appear).

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11), 4302-4311.

Kim, S. H., Kim, N. K., Ahn, S. C. & Kim, H. G. (1998) Object oriented face detection using range and color information. *Proceedings of the Third International Conference on Face and Gesture Recognition*, Nara, Japan.

Leung, T. K., Burl, M. C., & Perona, P. (1995). Finding faces in cluttered scenes using random labeled graph matching. *Proc. Intl. Conf. On Comp. Vis.,* 637-644.

Mondloch, C. J., Lewis, T. L., Budreau, D. R., Maurer, D., Dannemiller, J. L., Stephens, B. R., & Kleiner-Gathercoal, K. A. (1999). Face perception during infancy. *Psychological Science, 10,* 419-422.

Morrone, M. C., Burr, D. & Ross, J. (1983). Added noise restores recognizability of coarse quantized images. *Nature*, 305, 226-228.

Reisfeld, D. & Yeshurun, Y. (1992) Robust detection of facial features by generalized symmetry, In *Proceedings of the 11th Conference on Pattern Recognition*, 117-120.

Rikert, T. D., Jones, M. J. & Viola, P. (1999). A cluster based statistical model for object detection. *Procs. IEEE Conf. On Comp. Vis. And Pattern Recog.,* 1046-1053.

Rowley, H. A., Baluja, S., Kanade, T. (1995) Human face detection in visual scenes. CMU technical report# CS-95-158R. *Advances in Neural Information Processing Systems 8*, 1996, pp. 875 - 881

Saber, E. & Tekalp, A. (1996). Face detection and facial feature extraction using color, shape and symmetry based cost functions. *Proc. Intl. Conf. On Pattern Recog.,* 654-658.

Sadr, J., Mukherjee, S., Thoresz, K., and Sinha, P. (2001) Fidelty of Local Ordinal Encodings. In *Neural Information Processing Systems* (in press).

Sun, Q. B., Huang, W. M. and Wu, J. K. (1998) Face detection based on color and local symmetry information. *Proceedings of the Third International Conference on Face and Gesture Recognition*, Nara, Japan.

Sung, K. K., and Poggio, T. (1994) Example based learning for view-based human face detection, AI Laboratory memo # 1521, MIT.

Terrillon, J. C., David, M. & Akamatsu, S. (1998). Automatic detection of human faces in natural scene images by use of a skin-color model and of invariant moments. *Proceedings of the Third International Conference on Face and Gesture Recognition*, Nara, Japan.

Thornhill, R. and Gangestad, S. W. (1993) Human facial beauty: averageness, symmetry and parasite resistance. *Human Nature*, **4**, 237-269.

Thoresz, K. and Sinha, P. (2001) Qualitative representations for recognition. *Proceedings of the Annual Meeting of the Vision Sciences Society,* Florida.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., and Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive neuropsychology*, 17(1/2/3), 257-279.

Torralba, A., Sinha, P. (2001) Statistical context priming for object detection. *Int. Conf. On computer Vision*, IEEE Press, Los Alamitos, CA.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In: *Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001)*, IEEE Computer Society Press, Jauai, Hawaii, December 8-14, 2001, (to appear).

Yang, G. and Huang, T. S. (1994) Human face detection in a complex background. *Pattern Recognition*, 27(1) pp. 53-63.

Yokoyama, T., Yagi, Y. & Yachida, M. (1998). Facial contour extraction model. *Proceedings of the Third International Conference on Face and Gesture Recognition*, Nara, Japan.

Wu, H., Chen, Q. & Yachida, M. (1999). Face detection from color images using a fuzzy pattern matching method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21, 557-563.