Computational Modeling of Face Recognition Based on Psychophysical Experiments

Adrian Schwaninger[1,2], Christian Wallraven[1], and Heinrich H. Bülthoff[1]

[1]Max Planck Institute for Biological Cybernetics, Tübingen, Germany
[2]University of Zurich, Switzerland

Recent results from psychophysical studies are discussed which clearly show that face processing is not only holistic. Humans do encode face parts (component information) in addition to information about the spatial interrelationship of facial features (global configural information). Based on these findings we propose a computational architecture of face recognition, which implements a component and configural route for encoding and recognizing faces. Modeling results showed a striking similarity between human psychophysical data and the computational model. In addition, we could show that our framework is able to achieve good recognition performance even under large view rotations. Thus, our study is an example of how an interdisciplinary approach can provide a deeper understanding of cognitive processes and lead to further insights in human psychophysics as well as computer vision.

Keywords: face recognition, component and configural processing, computational modeling

Reliably recognizing faces is certainly one of the most important abilities of the human cognitive vision system. Bahrick, Bahrick, and Wittlinger (1975) reported that people could recognize familiar faces with an accuracy of 90 percent or more even when some of those faces have not been seen for fifty years. In contrast to faces, different object classes can often be distinguished based on relatively distinctive features like color, texture or global shape. Faces however form a quite homogenous stimulus category. In contrast to recognizing different object classes, face recognition relies on detecting subtle differences between facial parts and their spatial relations in order to recognize an exemplar within the category of faces. Such processing is very orientation-sensitive. When faces are turned upside-down they are much more difficult to recognize than it is the case for other objects (Yin, 1969; for reviews see Schwaninger, Carbon, & Leder, 2003; Valentine, 1988). Farah, Tanaka, and Drain (1995, p.633) answer the question "Why is face recognition so orientation sensitive?" in the following way: "Face perception is holistic and the perception of holistically represented complex patterns is orientation sensitive." According to their view, upright faces are stored as unparsed perceptual wholes in which individual parts (components) are not explicitly represented (see also Tanaka & Farah, 1991, Tanaka & Farah, 1993). The concept of holistic processing has been discussed controversially in the nineties. This might seem surprising if it is taken into account that a much earlier study could clearly show that face perception is based on different dimensions (Groner, 1967). Using multidimensional scaling techniques he could identify 12 different dimensions, most of which could be interpreted verbally (e.g., inter-eye distance, height of forehead, etc.). Based on many subsequent studies a distinction between component and configural information has been suggested, which provides an alternative hypothesis to purely holistic processing. We first review the empirical evidence for this view. We then present a model of face processing, which is able to integrate different theoretical concepts. Our computational implementation of this model is very consistent with results found in psychophysical experiments and achieves good recognition performance even under large view rotations, similar as humans.

Component and Configural Information

The term component information (or featural, piecemeal, part-based information) has been referred to facial elements, which are perceived as distinct parts of the whole such as the eyes, mouth, nose or chin. In contrast, the term configural information refers to the "spatial interrelationship of facial features" (Bruce, 1988, p. 38). Several studies have investigated the component configural hypothesis by replacing or altering component information (e.g., by replacing the nose or darkening teeth) or by changing configural information (e.g., by increasing the inter-eye distance). It was generally found that processing configural information is strongly impaired when faces are turned upside-down whereas processing component information remains relatively unaffected (e.g., Leder & Bruce, 2000; Murray, Yong, & Rhodes, 2000; Schwaninger & Mast, 1999; Searcy & Bartlett, 1996; Sergent, 1984; for a review see Schwaninger, Carbon, & Leder 2003). However, one possible caveat of these studies could be that selective manipulations of component and configural information are difficult to achieve. For example replacing the nose (component change) sometimes can alter the distance between the contours of the nose and the mouth which is a configural alteration. Similar difficulties apply to configural manipulations: For example moving the eyes apart (configural change) can lead to an increase of the bridge of the nose, which is a component change.

Problems like these can be avoided by selectively reducing component or configural information using scrambling and blurring procedures (e.g., Collishaw & Hole, 2000; Davidoff & Donnelly, 1990; Sergent, 1985). Schwaninger, Lobmaier, and Collishaw (2002) extended previous research by ensuring that scrambling and blurring effectively eliminate configural and component information separately. Furthermore, in contrast to previous studies, Schwaninger et al. (2002) used the same faces in separate experiments on unfamiliar and familiar face recognition to avoid potential confounds with familiarity. The results of two experiments are depicted in Figure 1, where the recognition performance is measured in d'-scores (Green & Swets, 1966; MacMillan & Creelman, 1991). In this study, d' was defined as the z-transform of the hit rate (percentage of correctly identified test images) minus the z-transform of the false alarm rate (percentage of images, which were erroneously identified as learnt images). In Experiment 1 it was found that previously learnt faces could be recognized by human participants even when the faces were scrambled into constituent parts (or components) so that configural information was eliminated (Figure 1, left). This result is consistent with the assumption of *explicit representations* of component information in visual memory. In a second condition, a low pass filter that made the scrambled part versions impossible to recognize was determined (Figure 1, middle). This filter was then applied to whole faces in order to create stimuli in which by definition local part-based information is eliminated or component and it can be tested whether configural information is explicitly encoded and stored. It was shown that configural versions of previously learnt faces could be

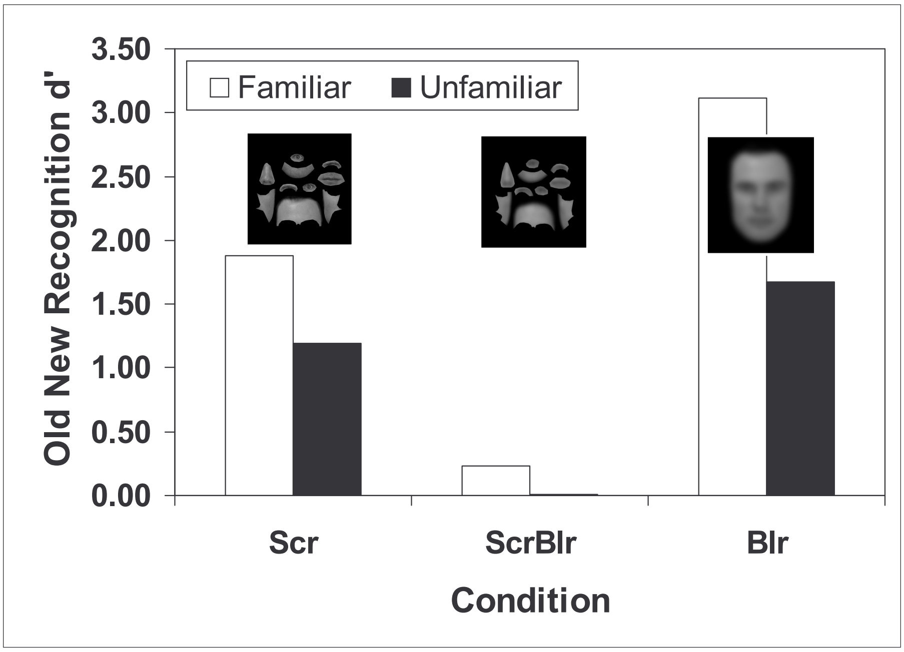

Figure 1. Results from Schwaninger et al. (2002). Previously learnt intact faces could still be recognized when they were scrambled into their components (Scr). Applying a low pass filter made the scrambled version impossible to recognize (ScrBlr). When the same filter was applied to whole faces recognition of these configural versions was well above chance (Blr). When the target faces were familiar (white bars), recognition performance increased but the relative balance between component and configural recognition remained the same.

recognized reliably (Figure 1, right), suggesting separate explicit representations of configural information. In Experiment 2 these results were replicated for subjects who knew the target faces (white bars in Figure 1). Both Experiments provided converging evidence in favor of the view that recognition of familiar and unfamiliar faces relies on separate visual representations of component and configural information.

Integrative Model of Face Recognition

Based on the results from different psychophysical studies, Schwaninger et al. (2002) and Schwaninger et al. (2003) have proposed the model depicted in Figure 2, which summarizes the cognitive architecture of face recognition. Processing entails extracting component information and global configural information in order to activate separate visual memory representations in higher visual areas (so-called face selective areas). The output of these representations converges to the same face identification units. These are holistic in the sense that they integrate the results of separate analyses of component and configural information. Adult face recognition is very orientation-sensitive because rotated faces overtax an orientation normalization mechanism (Rock, 1973, 1974, 1988; Valentine & Bruce, 1988; Schwaninger & Mast, 2004). It is not possible to mentally rotate a face as a whole and the component and configural representations that have been learnt based on exposure to upright faces can not be matched reliably.

Although faces become very difficult to recognize when turned upside down, changes in viewpoint are less detrimental to face recognition performance (e.g., Hill, Schyns, & Akamatsu, 1997; O´Toole, Edelman, & Bülthoff, 1998; Troje & Bülthoff, 1996; Wallraven, Schwaninger, Schumacher, & Bülthoff, 2002). When participants are familiar with faces they can recognize them even in 90° views (Bruce, 1982; Valentin, Abdi, & Edelman, 1997). If subjects are not familiar with a face, their view generalization capability remains relatively stable for view changes of about 15° but afterwards recognition performance drops substantially (Wallraven et al., 2002). Since generalization across viewpoint is one of currently investigated problems in computational vision of face recognition, it is certainly interesting to examine whether a model based on results from human psychophysics achieves better view generalization performance than other current approaches.

In summary, the computational modelling had the following objectives: 1) define the face representation, 2) model component and configural processing, 3) test the computational model using exactly the same stimuli as in the psychophysical experiments by Schwaninger et al. (2002), and 4) investigate recognition performance across viewpoint changes.

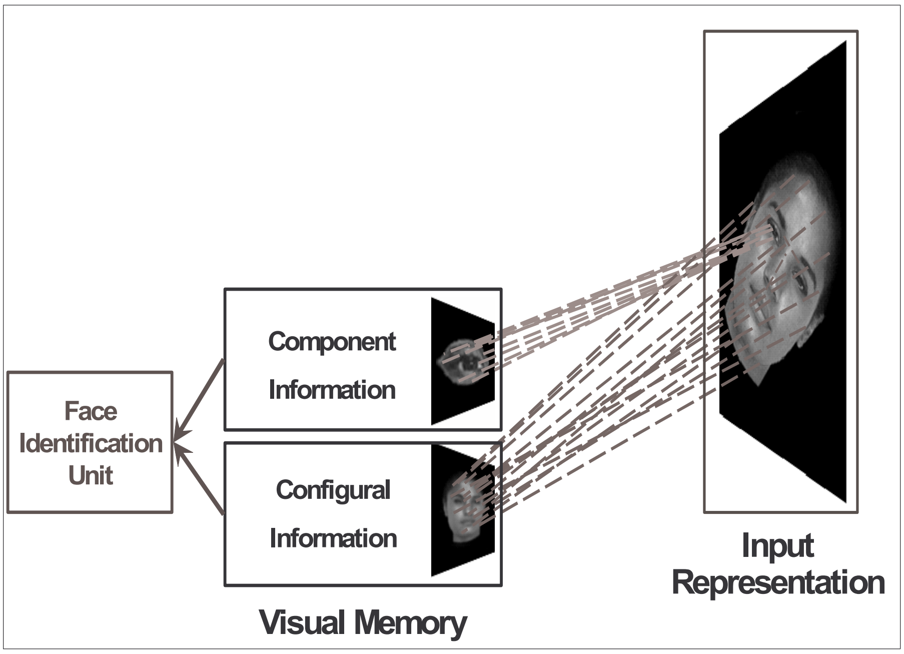

Figure 2. Integrative model of face processing after Schwaninger et al. (2002; 2003). From the input representation component and configural information is extracted in order to activate separate representations in visual memory. The outputs of these representations converge towards the same face identification units that integrate the component and configural inputs.

Computational Modeling

Face Representation

Our computational implementation in this paper is based on previous studies (Wallraven & Bülthoff, 2001a, 2001b) where an earlier version was successfully used to model psychophysical results on view-based object recognition (Wallraven et al. 2002).

The algorithm for constructing the face representation proceeds as follows: First, an input image is processed at two scales of a Gaussian pyramid to extract visual features, which form the basis of our face representation. Visual features are extracted by using an interest point detector (such as a standard corner detector, Wallraven & Bülthoff, 2001a), which yields pixel coordinates of salient image regions. Saliency here is defined at the pixel intensity level and can in our case be equated with regions that exhibit high curvatures of pixel intensities within their neighborhood. Around each of these located points we extract a small neighborhood of 5x5 pixels which captures local appearance information. From a computational point of view, this whole process of feature extraction not only reduces the amount of storage needed for the face representation significantly but also represents an efficient and more robust way of "image" processing. In addition, one can also motivate this choice of feature extraction from both psychological and physiological studies, which support the notion of visual features of intermediate complexity in higher brain areas (Ullman, Vidal-Naquet, & Sali, 2002).

In addition to these image fragments, we also determine for each feature its *embedding*, which consists of a vector containing pixel distances to a number of neighboring features. This vector of distances is used during the matching stage to determine either the component or the configural properties of each feature. In order to facilitate processing, the extracted distance vectors are sorted in increasing order. Figure 3 shows a reconstruction of a face from such a feature representation, in which features from coarse scales were resized according to the scale difference and then images from all scales superimposed starting with the coarsest scale.

Two aspects are worth noting here: First, even though our representation is quite sparse in terms of compressing the original data (a total of 160 features each of which contains 25 pixels results in a compression rate of 93.9%), the reconstruction still gives a fairly good visual impression of the original face. Second, and perhaps more importantly, one can see that the extracted features tend to cluster around important *facial* features. Eyes, mouth and nose are represented with a much higher density of features than, for example, the forehead or the cheeks. We certainly do not want to claim that a simple corner detector is able to explain the complicated processes of feature formation as there are of course other types of information available, of which facial motion resulting from facial expressions or talking is probably one of the most prominent. It seems, however, that our visual features already correspond at least in part to perceptually relevant *facial* features. This observation leads directly to our approach of defining component and configural representations. First of all, we do not want to use prior knowledge about facial components, which would for example be available in the form of state-of-the-art facial feature detectors (see Hjelmas & Low, 2001 for a recent overview) or the use of a sophisticated three-dimensional face model (such as the morphable model by Blanz and Vetter, 1999). Instead we opted for a purely bottom-up data-driven definition of such "components", which can accommodate different object classes but at the same time is flexible enough to allow later learning of a more abstract definition of parts. Components in our framework are thus defined as *tightly packed conglomerates of visual features at detailed scales.* This definition captures all the important aspects of the integrative model defined above without using prior knowledge in the form of pre-learned part models (such as templates for the eyes or spline models for the mouth). It should be noted, that components in our framework are defined by a form of *configuration* in the feature set. This implies that processing of components relies on the *relationship between features at detailed scales*. Complementing this component processing we can now define configural processing as the *relationship between features at coarse scales.* One important aspect of our definition of component processing is that it does *not* include an explicit clustering of the visual features into perceptual parts. Rather, the psychophysically found characteristics of component processing are made explicit only during *matching*, which will be outlined in the next section.

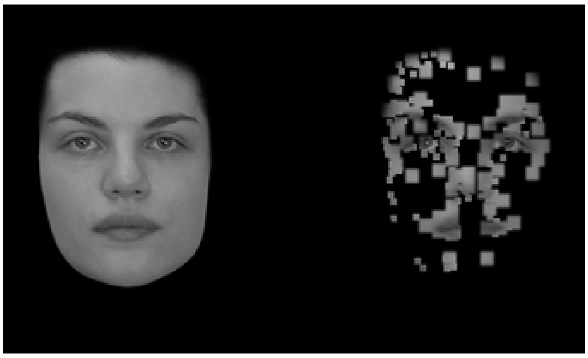

Figure 3. Left: original face. Right: Reconstruction from its feature representation. Blurred features originate from the coarse scale, whereas detailed features originate from the fine scale.

One view of component and configural processing would be that configural processing relies on the relationship between the extracted components. This would have an important consequence for the processing of faces, namely, that component detection comes first followed in a second step by configural processing. Equally valid, however, would be the assumption that the configural route is activated first by a coarse configural description of the stimulus, which then is able to trigger a more detailed processing of component information. The advantage of the second strategy lies in its coarse-to-fine processing, where information from configural processing can be used for later matching of components. As the psychophysical experiments using scrambled faces show (Schwaninger et al., 2002), humans seem to be able to match components *without* the help of configural processing with a reasonable performance. This means that although these two routes can be seen to rely on each other under 'normal' circumstances (i.e., for recognition of intact faces), the experimental evidence also speaks for two rather independent processing structures. In this paper, we will thus adopt the coarse-to-fine strategy as it closely follows our multi-scale representation, but we will keep the two types of processing largely separated. We have tried to address these issues in the design of the matching process, which is outlined in the following section.

Component and Configural Processing

The algorithm for recognition of face images is the second main part of the computational modeling. As each image consists of a set of visual features, recognition in our case amounts to finding the *best matching feature set* between a test image and all training images. The two routes for face processing are reflected by two types of matching algorithms based on configural and component information.

Matching of two feature sets is done by an algorithm inspired by Pilu (1997). First, a similarity matrix A is constructed between the two sets, where each term A_{ij} in the matrix is of the form of:

$$A_{ij} = \exp(-\frac{1}{\sigma_{app}^2} app^2(i,j)) \cdot \exp(-\frac{1}{\sigma_{emb}^2} emb^2(i,j)) \qquad (1)$$

The first term in equation (1) specifies the appearance similarity (*app*) of two visual features, whereas the second term determines the similarity between the embeddings of the features (*emb*). Appearance similarity is determined by the normalized grey value cross-correlation between the two pixel patches *i* and *j*, which was shown to give good results in previous studies (Wallraven & Bülthoff 2001a, 2001b; Wallraven et al. 2002). Embedding similarity is defined by the χ^2 difference between two distance histograms:

$$emb^2(i,j) = \chi^2(i,j) = \frac{\sum_{k\in N}(d_i(k) - d_j(k))^2}{\sum_{k\in N} d_i(k) + d_j(k)} \qquad (2)$$

where *d(k)* is the vector containing distances to all other features sorted in increasing order. *Component* matching is done in our framework by restricting *N* to the first few elements of the histogram, thus restricting analysis to *close* conglomerates of features – a *local* analysis. *Configural* matching on the other hand relies on *global* relationships, such that *N* is restricted to the last elements of the sorted distance histogram. The size of N should be small for component matching (in our experiments, we used $|N|=3$) and larger for the global configural matching (in our experiments, we used $|N|=|d|$). In addition, the parameters σ_{app} and σ_{emb} can be used to control the relative importance of the two types of information.

The matrix **A** thus captures similarity between two feature sets based on a combination of distance information and appearance information. Corresponding features can now be found with a greedy strategy by looking at the largest elements of **A** both in row and column satisfying *A(i,j)>thresh* (Pilu, 1997; Wallraven & Bülthoff 2001a, 2001b), which yields a one-to-one mapping of one feature set onto the other. The threshold *thresh* is used to introduce a quality metric for the matches. The percentage of matches between the two feature sets for the component route *and* the configural route then constitute the two final matching scores.

Modeling Psychophysical Experiments

In this section, we summarize the experiments in which the proposed framework was applied to the stimuli of the psychophysical experiments explained above (see Figure 1). In order to increase the statistical power, all experiments were carried out on *10 different splits* of the database of 20 faces into 10 targets and 10 distractors. In addition, we have calculated d'-scores for the same set of stimuli to facilitate comparison with human data. Figure 4 shows the results of the computational modeling experiments with separate bars for component and configural processing.

In agreement with the psychophysical results (compare Figure 1 and Figure 4), we found that scrambled and blurred stimuli could be recognized with a significant advantage of configural over component processing. As was to be expected, both types of processing break down in the scrambled-blurred condition, where the information could support neither detailed component nor global configural analysis. These results demonstrate that

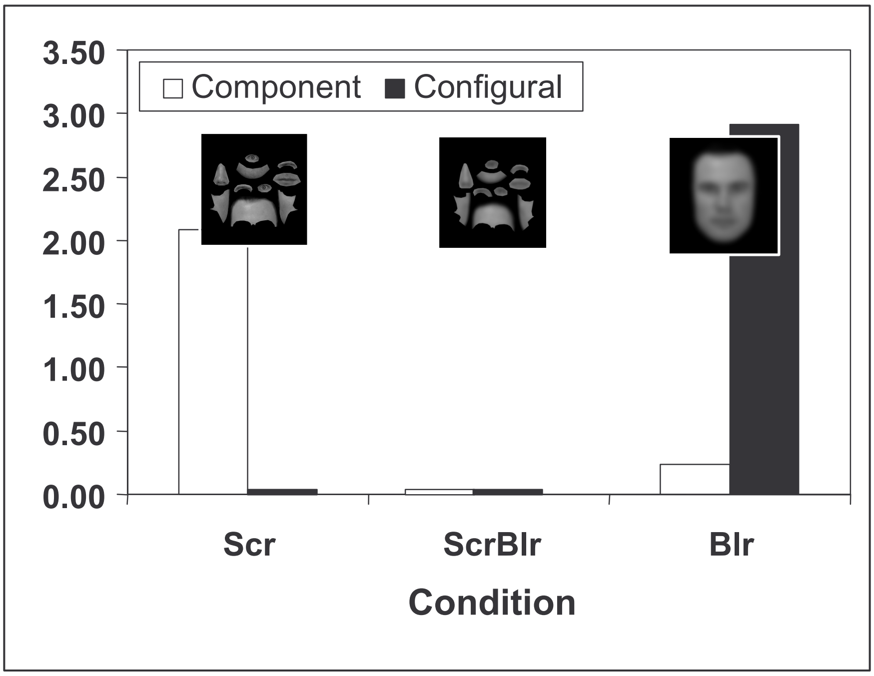

Figure 4. Computational modeling results for the psychophysical experiments explained in the *Component and configural information* section (compare with Figure 1). Shown is the output of both the component and the configural route as d' values. Note how the different routes are active in the different conditions.

our framework is able to capture the characteristics of the two separate routes as found in the psychophysical experiments.

Figure 5 shows an example of the scrambled and the blurred condition for the two types of matching, where corresponding features are indicated as white dots. For better visual analysis, we have shown just the component processing for the scrambled condition and the configural processing for the blurred condition (there are only a few other matches in either condition from the other route). One can see that component matching concentrates on high-level details such as corners of the mouth, points on the nose, some features in the eyes and on the eyebrow, etc. (disregarding points introduced by the cutting procedure for the scrambling or by the outline of the face; these points are treated as artifacts and produce only a *constant* effect for all stimulus conditions). Interestingly, this observation directly leads to concrete experimental predictions, which can be used to design further psychophysical studies: most of the matching features in the component processing route are in high-contrast regions (due to the nature of our visual features). If component processing in humans relies on similar low-level information, components in this experiment, which have less high-contrast regions (such as the forehead or the cheeks) should contribute *less* to the human recognition score. We are currently designing a set of experiments, which directly address this question of how different parts are weighted. This represents a good example of how computational modeling feeds back into cognitive research. Going back to Figure 5, configural matching on the other hand results in a more global match of the face with corresponding features spread evenly over the whole face area. Interestingly, this property also allows for categorization of faces, as the configural information captures the global layout of face structure and can thus be used for a coarse labeling of faces versus non-faces. Figure 6 shows an example of the matching result between two different faces, which demonstrates the generalization capabilities of our framework. In this case, the configural route is activated most strongly (in a test with 20 faces, on average 95% of the matches) showing that class-based information is mainly

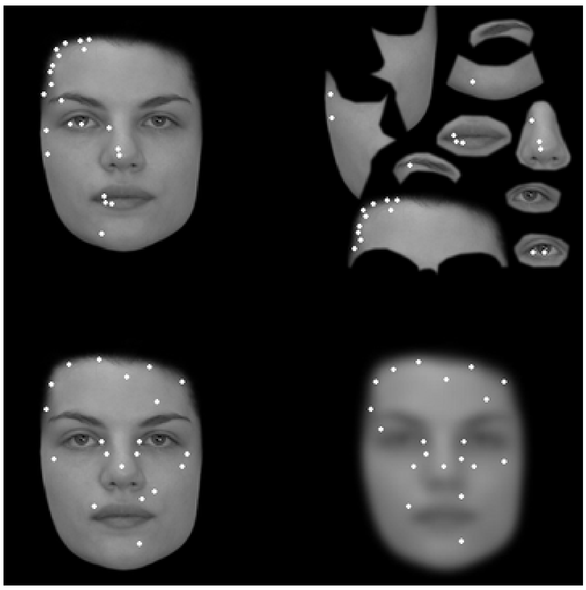

Figure 5. Corresponding features for the two test conditions (upper row: scrambled, all correspondences from part-based route, lower row: blurred, all correspondences from configural route).

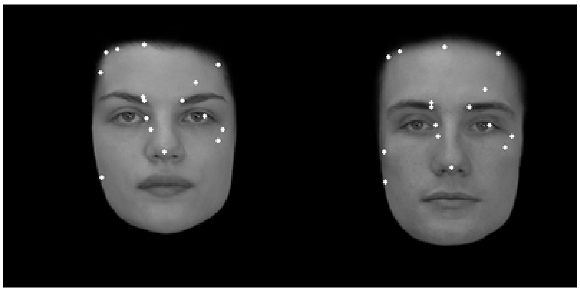

Figure 6. Corresponding features for two faces showing that the general class-based similarity in the layout of facial features is well captured by the configural information of our system.

captured by global feature layout for our chosen representation.

Recognition under View Rotations

In this section, we want to demonstrate that apart from being able to model psychophysical experiments, our framework also offers significant performance improvements in other application areas. Here, we report preliminary results for recognition of faces under view rotations.

It is well-known that finding corresponding features between two images, which show the same face (or object) under a large depth rotation, is a very difficult problem. Apart from the fact that there is of course an upper bound to invariant extraction of features under such large rotations, the issue of false matches becomes more prominent as feature similarities start to decrease with increasing distance to the test view. Whereas it is difficult to find a large number of consistent matches across large pose differences, one type of information that can be used as an additional constraint is the *configuration* of the features. After all, some aspects of the distance relationships between features will stay constant during a depth rotation.

Figure 7 shows two matching results for two views of a face, which differ by 30 degrees. The first row shows corresponding features for the matching procedure proposed above using configural and component matching. In total, 3% of the features could be matched for the component route (one false match, shown as a circled dot in the left figure), whereas 32% could be matched for the configural route (no false matches). This result demonstrates that global configural information survives even larger view rotations without introducing too many false matches. For comparison, the bottom row shows the same faces with a matching algorithm used in Wallraven and Bülthoff (2001a, 2001b) and Wallraven et al. (2002), where matching relies on a combination of simpler distance and appearance constraints; in particular the embedding measure is replaced by a simple Euclidean pixel distance measure:

$$A_{ij} = \exp\left(-\frac{1}{\sigma^2_{app}} app^2(i,j)\right) \cdot \exp\left(-\frac{1}{\sigma^2_{emb}} dist^2(i,j)\right) \qquad (3)$$

This method not only produces less matches (18% in total), but also more false matches (5 false matches indicated as circled dots in the left face) showing that a simple distance constraint is not robust enough to survive large view rotations. Recognition results with a database of 30 faces show that our algorithm is able to recognize faces under *30 degrees* view rotation with a recognition rate of 100%, whereas the standard matching procedure yields only 57% recognition rate. This demonstrates that the use of local features not only reduces storage requirements but also greatly enhances recognition capabilities. In addition, our results show that the use of configural and component processing yields an excellent generalization performance across views.

The current limit at which the configural matching procedure still performs at 100% for this database – using the same set of matching parameters - lies at *36 degrees*. This result could be even further improved by incorporating class-specific prior knowledge (see Blanz, Romdhani, & Vetter, 2002 for the most

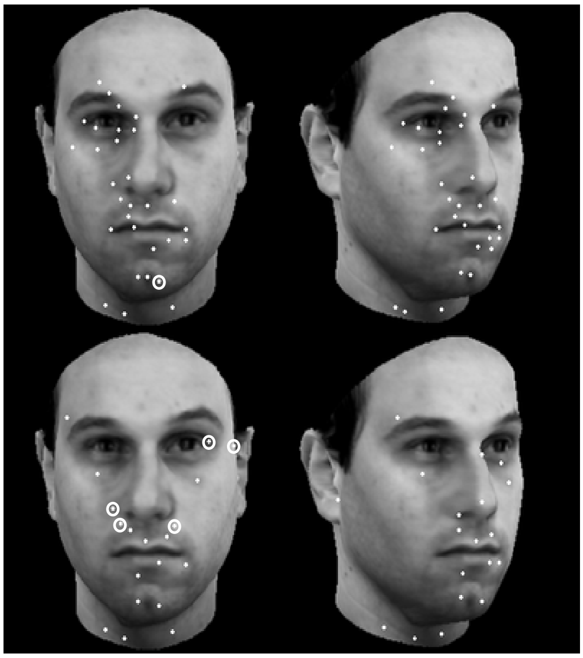

Figure 7. Matching under view rotations of 30°. Upper row: corresponding feature locations found using the configural/component matching method. Lower row: corresponding features using a standard matching method. Note, that the number of false matches (shown as circled dots in both rows) increases for the standard method. (Stimuli taken from the MPI database of 200 three-dimensional laser-scanned faces, http://faces.kyb.tuebingen.mpg.de)

extreme case of applying prior knowledge with the help of a detailed three-dimensional morphable face model).

In addition to class-based knowledge there are additional constraints one can use for increasing the number and accuracy of matches even further. As was shown in Wallraven et al. (2002), the knowledge about which kind of rotation occurs between two images can be used to further enhance matching performance by biasing the matching direction towards the direction of rotation. We are currently conducting further experiments with this extended approach.

Nevertheless, we want to stress that the amount of view generalization we have demonstrated here already represents a promising step towards human performance levels (e.g., Hill, Schyns, & Akamatsu, 1997; O´Toole, Edelman, & Bülthoff, 1998; Troje & Bülthoff, 1996; Wallraven, Schwaninger, Schumacher, & Bülthoff, 2002).

Conclusions

Psychophysical evidence strongly supports the notion that face processing relies on two different routes, which are represented by configural information and component information. We have implemented a computational model of such a processing architecture, which is based on this separation of the two types of information. Visual features extracted from interest points form the basis of our proposed representation. Configural and component processing were implemented based on the distance histograms between feature locations. In a second step, we have tested our architecture using stimuli from psychophysical experiments in order to examine its performance and modeling capabilities. The results of this experiment were very similar on a qualitative level to human performance. In this context it has to be said that an exact quantitative modeling – while this might seem a desirable goal – cannot be realistically achieved as there are too many hidden variables in the exact formation of the psychophysical data. A qualitative similarity on the other hand is a sign that the basic assumptions of the computational architecture and its implementation share a similar structure. This argumentation leads to the second important step in cognitive modeling: closing the loop between psychophysics and modeling. Our computational experiments have shown for example that certain features seem to have a higher saliency than others. In this case, a psychophysical experiment which validates these findings will give even stronger support for the proposed architecture.

Furthermore, the psychophysical data supports also the notion that global processing is very orientation-sensitive (orientation in the picture plane), whereas local component processing is not. We are currently running experiments with our proposed computational architecture in order to simulate the face inversion effect.

Finally, we have shown results from recent experiments, which show that our framework leads to significant performance improvements for recognition of faces under large view rotations. As a next step, we will conduct large scale experiments in order to compare our approach against current state-of-the-art algorithms. Our results, however, already demonstrate the usefulness of integrating results from cognitive research into computational systems in order to take a step towards the level of recognition performance demonstrated every day by the human visual system.

References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54-75.

Blanz, V., Romdhani, S., & Vetter, T. (2002). Face identification across different poses and illuminations with a 3D morphable model. In *Proc. 5th Int. Conference on Automatic Face and Gesture Recognition*, 202-207.

Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3D faces. *In Proc. Siggraph99,* 187-194.

Bruce, V. (1982). Changing faces: Visual and non-visual coding process in face recognition. *British Journal of Psychology, 73,* 105-116.

Bruce, V. (1988). *Recognising faces*. Hillsdale, NJ: Lawrence Erlbaum.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception, 29*, 893-910.

Davidoff, J., & Donnelly, N. (1990). Object superiority: A comparison of complete and part probes, *Acta Psychologica, 73*, 225–243.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance, 21*, 628-634.

Green, D. M. & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Groner, R. (1967). Dimensionen der Wahrnehmung von Gesichtern [Dimensions of face perception]. *Zeitschrift für experimentelle und angewandte Psychologie, XIV*(1)*,* 135-154.

Hill, H., Schyns, P. G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition, 62,* 201-222.

Hjelmas, E., & Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding. 83,* 236-274.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology, 53A*, 513-536.

McMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* New York: Cambridge University Press.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside-down faces. *Psychological Science, 11*, 498-502.

O´Toole, A., Edelman, S., & Buelthoff, H. H. (1998). Stimulus-specific effects in face recognition over changes in viewpoint. *Vision Research, 38,* 2351-2363.

Pilu, M. (1997). A direct method for stereo correspondence based on singular value decomposition. *Proc. Computer Vision and Pattern Recognition*, 261-266.

Rock, I. (1973). *Orientation and form*. New York: Academic Press.

Rock, I. (1974). The perception of disoriented figures. *Scientific American, 230*, 78-85.

Rock, I. (1988). On Thompson's inverted-face phenomenon (Research Note). *Perception, 17*, 815-817.

Schwaninger, A., Carbon, C. Ch., & Leder, H. (2003). Expert face processing: Specialization and constraints. In G. Schwarzer & H. Leder, *The development of face processing* (pp. 81-97). Seattle: Hogrefe & Huber.

Schwaninger, A., Lobmaier, J., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science, 2525,* 643-650.

Schwaninger, A., & Mast, F. (1999). Why is face recognition so orientation-sensitive? Psychophysical evidence for an integrative model. *Perception (Suppl.), 28*, 116.

Schwaninger, A. & Mast, F. (2004). *The face inversion effect can be explained by capacity limitations of an orientation normalization mechanism*. Manuscript submitted for publication.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information of faces. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 904-915.

Sergent, J. (1984). An investigation into component and configurational processes underlying face recognition. *British Journal of Psychology , 75,* 221-242.

Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*(6), 846-61.

Tanaka, J. W. & Farah, M. J. (1991). Second-order relational properties and the inversion-effect: Testing a theory of face perception. *Perception & Psychophysics, 50,* 367-372.

Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*, 225-245.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research, 36*(12), 1761-1771.

Ullman, S., Vidal-Naquet, M., & Sali., E. (2002) Visual features of intermediate complexity and their use in classification. *Nature Neuroscience, 5*(7)*,* 682-687.

Valentin, D., Abdi, H., & Edelman, B. (1997). What represents a face: A computational approach for the integration of physiological and psychological data. *Perception, 26*, 1271-1288.

Valentine, T. (1988). Upside-down faces: a review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.

Valentine, T., & Bruce, V. (1988). Mental rotation of faces. *Memory and Cognition, 16,* 556-566.

Wallraven, C., & Bülthoff, H. H. (2001a). Automatic acquisition of exemplar-based representations for recognition from image sequences. Workshop on Models vs. Exemplars. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*.

Wallraven, C., & Bülthoff, H. H. (2001b). View-based recognition under illumination changes using local features. Workshop on Identifying Objects Across Variations in Lighting: Psychophysics and Computation. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*.

Wallraven, C., Schwaninger, A., Schuhmacher, S., & Bülthoff, H. H. (2002). View-based recognition of faces in man and machine: Re-visiting inter-extra-ortho. *Lecture Notes in Computer Science, 2525,* 651-660.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Author Note

This work was supported by a grant from the European Commission (CogVis, IST-2000-29375).

Adrian Schwaninger

Max Planck Institute for Biological Cybernetics
Dept. Bülthoff
Spemannstr. 38
72076 Tübingen, Germany
E-Mail: adrian.schwaninger@tuebingen.mpg.de.