Becoming *Visually* Familiar

M. Castrillón-Santana, O. Déniz-Suárez,
J. Lorenzo-Navarro and D. Hernández-Sosa
IUSIANI
Edif. Ctral. del Parque Científico Tecnológico
Universidad de Las Palmas de Gran Canaria
Las Palmas de Gran Canaria, 35017
Spain
mcastrillon@iusiani.ulpgc.es

Abstract

Automatic face recognition has been mainly tackled by matching a new image to a set of previously computed identity models. The literature describes approximations where those identity models are based on a single sample or a set of them. However, face representation keeps being a topic of great debate in the psychology literature, with some recent results suggesting the use of an average image. In this paper, instead of restricting our system to a fixed and precomputed classifier, the system learns iteratively based on the experience extracted from each meeting. The experiments presented introduce the use of an exemplar average based approach. The results show similar performance to an approach based on the use of multiple exemplars per identity, but reducing storage and processing cost. The process is done autonomously, using an automatic face detection system that meets people, excepting the supervision provided by a human to confirm or correct each meeting classification suggested by the system.

1 Introduction

It is known that the human and primates have an ability to distinguish a large amount of individuals under very different conditions [31]. However, this effortless capability keeps being a topic of debate in the psychology community, particularly the way faces are represented in the brain. Different experiments have in the past suggested that an object model requires a collection of images [36], while others indicate that an average of them is used [7, 31]. In both cases, these exemplars are actively collected along the human visual system operation, i.e. are provided by experience.

Automatic face recognition is being tackled differently. Automatic face recognizers are typically built in batch mode, instead of assuming the iterative process carried out by humans until high reliability is achieved by young adults [23]. Automatic systems do not change their model even when faces change throughout life. For those automatic systems, their performance, for a given database, is measured in terms of recognition rate [26, 27]. However, even when brilliant performances have been achieved in that scenario, being in cases even better than humans [7, 8, 20, 28], these systems are still far from being comparable to humans in most real life situations [1, 7]. This difference is particularly notorious in the context of familiar faces, where the human system evidences an impressive performance [7, 8]. This latter scenario is basically our aim, even when automatic face recognizers commonly avoid it.

Evolution is needed in such a context because it is not proven that the results achieved with a training database can be extended to the whole face domain. This is the essential idea behind Wolpert's No Free Lunch Theorem [37], which states that, on the criterion of prediction performance, there are no reasons to prefer the hypotheses selected by one learning algorithm over those of another. The perfect fit to a training set does not guarantee low error for future, unseen samples.

As mentioned above, in this paper we focus on the problem of online building a representation for a set of familiar faces using the information provided by a real time automatic face detector. For that purpose, we employ a common face recognition algorithm for face representation and classification, with the restriction that it must be suitable for real time performance.

In order to confirm if successive expositions to a set of identities improves the face recognition performance, i.e. becoming familiar, a set of video streams have been gathered. Later, we expose randomly a set of identities several

times to the system and analyze if their recognition error rate decreases as more experience is gained by the system. As stated above, the system classification mechanism will not be fixed, but it will evolve based on those expositions, i.e. on system experience.

The paper describes first the mechanism for detecting faces and selecting samples from video. The representation and classification approaches are briefly presented later. Finally the experiments and conclusions are outlined.

2 Automatic face detection

Cue combination provides greater speed and robustness in the face detection problem [10]. Thus, their performance outperforms single cue based detection such as [35], providing a more reliable tool for real time interaction. The availability of different approaches for research purposes, as for example [10], provides the community a source for video processing. Briefly, this system combines two different Viola-Jones' framework based detectors [22, 35], both available in the OpenCV library [18], to initially detect a face, which is then modelled and later redetected using a combination of face and facial elements tracking, and face and torso color based techniques.

2.1 Facial elements detection

More precise facial feature detection helps reducing the influence of the misalignment introduced by automatic face detectors. Using the information provided by the face detector described in [10], we have integrated additional facial features detection (eyes, mouth and nose) based on the general object detection framework by Viola and Jones [34]. Positive samples were obtained annotating by hand the eyes, nose and mouth location in 7000 facial images taken randomly from Internet and selected samples from facial databases such as BIOID [14]. The images were later normalized by means of eye information to 59×65 pixels, see Figure 1 for a normalized image example. Five different detectors were then computed: 1-2) Left and right eye (18×12 pixels), 3) eye pair (22×5), 4) nose (22×15), and 5) mouth (22×15). All these Viola-Jones based classifiers have been made available to the OpenCV community [29].

The facial feature detection procedure is applied only in those areas which present evidence of containing a face. This situation happens in those areas where a face has been detected in the current frame, or in those where a face was detected in the previous frame (video processing). Given the estimated area for each feature, considering a frontal pose, candidates are searched both by means of those Viola-Jones' detectors and by tracking [15] (only for video processing). Once all the candidates have been obtained, the best combination is selected choosing the one with the highest likelihood based on the normalized positions for eyes, nose and mouth, see Figure 1.

Figure 1. Normalized face sample and likely locations for nose tip and mouth center positions after normalization. White areas reflect most likely locations.

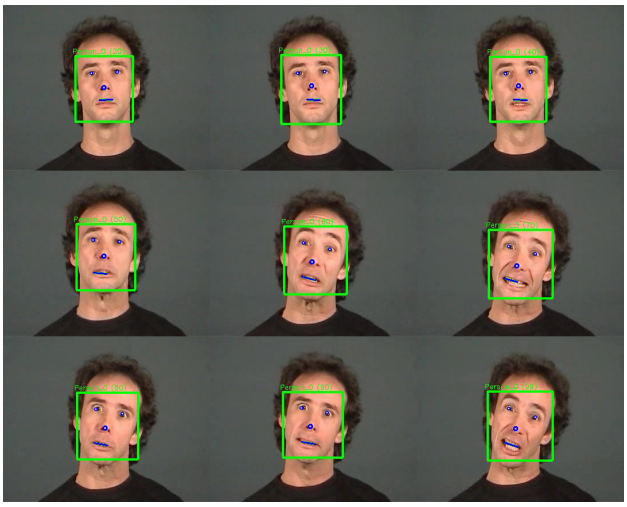

Figure 2. Facial element detection samples for a sequence extracted from DaFEx [4].

Figure 2 presents some results achieved with a sequence extracted from DaFEx [4], that presents changes in expression and in plane rotation, evidencing the possibilities of this facial features approach.

3 Face representation and classification

Face images have a high dimensionality. Principal Components Analysis (PCA) decomposition [21, 32] has been frequently applied to represent them in a lower dimensionality space. An image is projected in the PCA space, and its coefficients are used instead, see Figure 3. In the experiments presented below, the first 70 coefficients were employed observing the results achieved in previous experiments using this representation space [11].

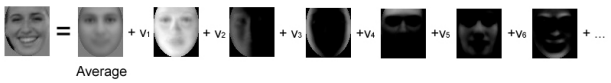

Figure 3. PCA decomposition

We assume that our system is initially unable to recognize faces, but it knows the face appearance as it is able to detect them. Therefore, an initial representation space is available, but it evolves later according to experience. This perspective would fit with the initial attraction of face patterns for newborns [30], and the later difficulties to recognize faces not normally seen in our environment, i.e. the *other race effect* [6]. Incremental PCA [16] is a mechanism that can adapt the representation space to the exposition to new patterns.

PCA is a common representation space used for face processing, being known its fast performance. Its problems with illumination are known in the literature [5], and they have prompted the apparition of alternative representation spaces such as Independent Components Analysis (ICA) [3], or Gabor filters [12, 25]. However, according to the results achieved in [13] that suggest that the selection of a powerful classification criterium is more critical than the representation space (PCA or ICA), recognition experiments have been carried out using Support Vector Machines (SVMs) [33].

The dataset used to compute the initial face representation space contains 7000 face images taken from Internet and selected samples from facial databases such as BIOID [14]. These images have been normalized according to their annotated eye positions obtaining 59×65 samples. In the experiments we assume two different uses of this PCA space: 1) keeping it fixed, and 2) employing an incremental PCA approach to improve the representation space with new experience [2, 16].

3.1 Sample selection from automatic face detection

The face detection approach employed, introduced briefly in Section 2, provides a normalized face anytime its eyes are detected. As mentioned above, we have additionally integrated nose and mouth detection based on the Viola-Jones' framework [35], achieving a maximal number of four facial features. For each individual detected, a set of normalized faces are selected from the whole collection of normalized faces extracted during the interaction session held with an individual. These selected images are the exemplars.

Exemplars are selected as follows: The first time that three of the facial elements are located, an exemplar is created. For a video stream, if at least three of those facial features are tracked in the next frame, no new exemplar will be created. Instead its persistence, pe, will be increased. At the end the exemplar persistence indicates the number of consecutive frames that it was tracked. A new exemplar will be created once the facial elements are lost, and their detection is newly performed, see Figure 4 for a graphical overview. Additionally, the PCA reconstruction error [17] is computed for each exemplar. This is done to estimate the *faceness* of the exemplar, and therefore reduce the use of misaligned patterns provided by the detector.

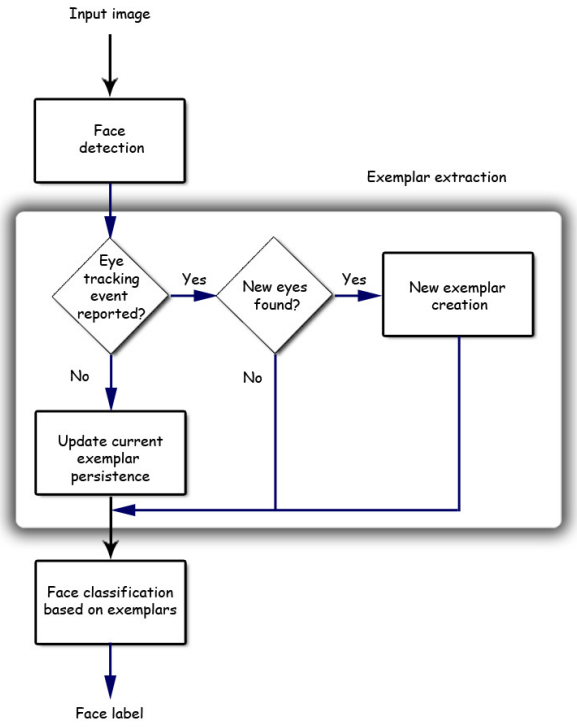

Figure 4. Exemplar selection mechanism.

An individual whose face has been detected in the current frame, is therefore represented by the collection of s exemplars extracted during the continuous interaction session or meeting. These exemplars, e_j, will be used to label the whole meeting or detection thread, d_t, the set of continuous detections for a face. This is done weighting each face-like exemplar by its persistence pe_j. The exemplar *faceness* is estimated selecting only those exemplars whose PCA reconstruction error is not notoriously bigger than the average obtained for the whole collection of exemplars.

$$P(C_k|dt) = \frac{\sum_{j=1}^{s} P(C_k|e_j) * pe_j}{\sum_{n=1}^{s} pe_n} \quad (1)$$

The favorite class, C_k, will be assigned to a face in the

current frame. At this point the system will be supervised. If the system was corrected, and the correct class was C_c, all the incorrectly labelled exemplars, i. e. $P(C_c|e_j) = 0$, will be added to the training set. If the supervisor confirmed the class suggested by the system, C_k, similarly incorrectly assigned exemplars, $P(C_k|e_j) = 0$, will be added to the training set.

4 Experiments

As stated above, our aim is to check if the continuous exposition of a face detection and recognition system to a set of identities allows the system to improve its models, and therefore its performance. To simulate this, while offering the ability to reproduce the experiment, we have gathered a database of sequences corresponding to 80 identities. We first checked the availability of a database containing videos of individuals recorded during a large period of time, i.e. days, weeks, or months. A database such as XM2VTS [24], recorded using the same camera, with a controlled illumination and background, is not well suited to verify the unrestricted problem tackled in this paper. For that reason we have built up a database making use of our own recordings during some months. The only restriction imposed for each recording is the presence of a single person. Therefore, these sequences were recorded using different cameras without controlled conditions.

The database contains 310 different video streams (320 by 240 pixels or larger) corresponding to those 80 identities. It must be noticed that the identities contained in this database are completely independent from those used to compute the initial PCA space, see Section 3. For each identity at least two sequences are available in this dataset.

This dataset have already been used in [9]. However, in that paper the classification was carried out by a serialized recognition and verification approach. Additionally, identities were exclusively modelled by means of multiple exemplars. In the experiments presented below, we will apply not only an approach based on multiple exemplars, but also an approach based on modelling each identity by the average of those exemplars. This is done with the purpose of verifying the importance recently given to the average image by the Psychology community [7, 19, 31].

For the first experiment only the subset of identities that have at least five sequences, i.e. meetings, available was processed. Figure 5 presents the results achieved for this subset of 24 identities. The experiment was performed three times meeting exactly five times each individual. For each run a randomly selected order was used, therefore the results are averaged. The reader can observe that the lowest error rate achieved take place after around 30 meetings. However, it must be noticed that this effect is produced by the randomness inherent in the experiments. After 30 meetings it is highly unlikely that all the identities, 24, have already been seen, therefore the recognition problem is easier. For this reason we should observe the evolution later, when it is likely that almost all the identities have already been met at least once. In the final stages, when all the identities have been repeatedly met, both incremental based approaches present a decreasing tendency in the recognition error.

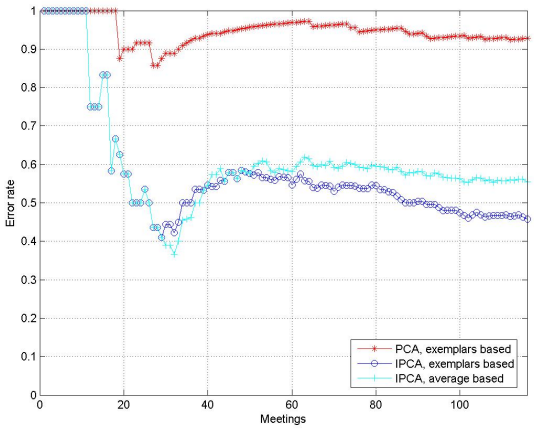

Figure 5. Accumulated error rate evolution for an experiment with 24 different identities and five sequences per identity. The results were averaged over three random runs.

The average exemplar based approach performs slightly worse. However, it presents a clear advantage in relation to the exemplar based approach. Indeed, retraining the classifier for an average based system is simpler in terms of computational cost, as only one sample is needed per class (for this experiment 174 exemplars are needed by the exemplar based approach vs. 24 samples by the average based approach). This fact makes the system suitable for online learning.

Results achieved in a more demanding experiment in terms of number of identities, are presented in Figure 6. In this experiment the total set of sequences available, i.e. 310 sequences of 80 individuals, was processed. For the second experiment it must be noticed that the number of sequences per identity is not homogeneous, i.e. some have only two available while others have up to ten. Obviously the latter provide more information to model their identity, according to our assumption, to build a better model, i.e. to become familiar.

Results suggest a slightly worse performance than in the first experiment. However, it must be reminded the higher difficulty of this problem, in terms of the number of identi-

ties and the average number of meetings per individual. In any case, the Figure evidences again that both Incremental PCA based approaches perform clearly better than the fixed PCA approach. Newly the number of samples used by the average based approach at the end of the experiment, 80 (identical to the number of identities), is clearly lower than the number achieved by the exemplars based approach, 469.

Further work should analyze if the automatic face detector can present some misalignment problems which could affect the average computation, see Figure 7 to observe the precision provided by the face detector for the selected exemplars of two identities.

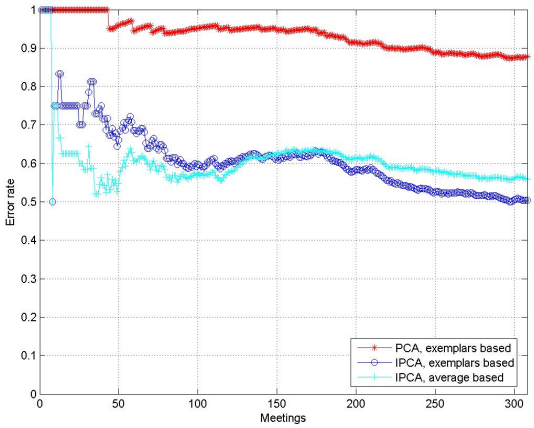

Figure 6. Accumulated error rate evolution for an experiment with 80 different identities and an irregular number of sequences per identity. The results were averaged over three random runs.

For both experiments the overall system behavior seems to be similar, i.e. incremental PCA provides a better tool to model those identities met, making them each time more familiar as the accumulated error rate decreases.

5 Conclusions

In this paper we have examined two different approaches for face representation in a real world face processing problem. An automatic system is repeatedly introduced to different identities. The information provided by the system from each meeting is later employed to initiate or update the model of that identity. Results evidence first that, if incremental PCA is used, these identities become more familiar as the number of meetings increases. It is also evidenced that the use of an average exemplar based approach provides slightly worse, but quite close, performance, while

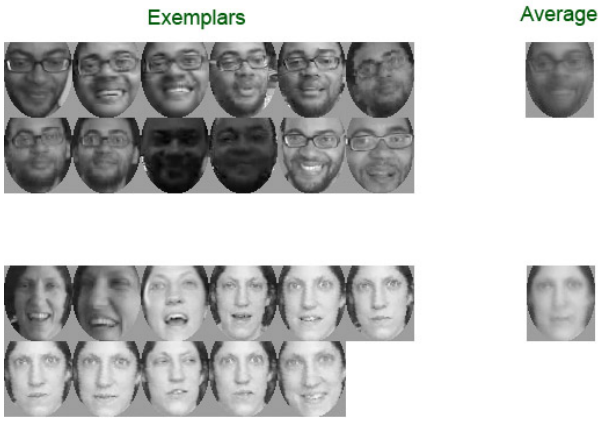

Figure 7. Exemplars and average image obtained for two sample identities modelled in the second experiment.

simplifying the training process, and reducing the storage requirements. The use of a face recognizer based on average images which is updated iteratively based on its experience could be suitable to assume the face evolution depending on age, as the model stored will be constantly updated.

These results fit with evidences observed for the average image in the human system [19]. Further work must confirm the promising results achieved in this paper with a larger experimental setup. A larger collection of sequences per identity will be collected in order to investigate if their model can keep on being improved.

Acknowledgments

Work partially funded by research projects Univ. of Las Palmas de Gran Canaria UNI2005/18 and the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

References

[1] A. Adler and J. Maclean. Performance comparison of human and automatic face recognition. In *Biometrics Symposium*, 2004.

[2] M. Artac, M. Jogan, and A. Leonardis. Incremental PCA for on-line visual learning and recognition. In *Proceedings 16th International Conference on Pattern Recognition*, pages 781–784, 2002.

[3] M. Bartlett and T. Sejnowski. Independent component of face images: a representation for face recognition. In *Procs. of the Annual Joint Symposium on Neural Computation,Pasadena, CA*, May 1997.

[4] A. Battocchi and F. Pianesi. Dafex: Un database di espressioni facciali dinamiche. In *SLI-GSCP Workshop Comunicazione Parlata e Manifestazione delle Emozioni*, Dezembre 2004.

[5] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. on PAMI*, 19(7):711–720, 1997.

[6] V. Bruce and A. Young. *The eye of the beholder*. Oxford University Press, 1998.

[7] A. M. Burton, R. Jenkins, P. J. Hancock, and D. White. Robust representations for face recognition: The power of averages. *Cognitive Psychology*, 51(3):256–284, 2005.

[8] A. M. Burton, S. Wilson, M. Cowan, and V. Bruce. Face recognition in poor quality video: Evidence from security surveillance. *Psychological Science*, 10(3), 1999.

[9] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. Learning to recognize faces by successive meetings. *Journal of Multimedia*, pages 1–8, 2006.

[10] M. Castrillón Santana, O. Déniz Suárez, M. Hernández Tejera, and C. Guerra Artal. ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, pages 130–140, April 2007.

[11] M. Castrillón Santana, J. Lorenzo Navarro, D. Hernández Sosa, and Y. Rodríguez-Domínguez. An analysis of facial description in static images and video streams. In *2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.

[12] J. Daugman. Complete discrete 2-d gabor transforms by neural networks for image analysis and compression. *IEEE Trans. on Acoustics, Speech, and Signal Processing*, 36(7), July 1988.

[13] O. Déniz Suárez, M. Castrillón Santana, and F. M. Hernández Tejera. Face recognition using independent component analysis and support vector machines. *Pattern Recognition Letters*, 24(13):2153–2157, September 2003.

[14] R. W. Frischholz and U. Dieckmann. Bioid: A multimodal biometric identification system. *IEEE Computer*, 33(2), February 2000.

[15] C. Guerra, M. Hernández, A. Domínguez, and D. Hernández. A new approach to the template update problem. In *2nd Iberian Conference on Pattern Recognition and Image Analysis*, Estoril, Portugal, June 2005.

[16] P. Hall, D. Marshall, and R. Martin. Incremental eigenanalysis for classification. In *British Machine Vision Conference*, volume 1, pages 286–295, September 1998.

[17] E. Hjelmas and I. Farup. Experimental comparison of face/non-face classifiers. In *Procs. of the Third International Conference on Audio- and Video-Based Person Authentication. Lecture Notes in Computer Science 2091*, pages 65–70, 2001.

[18] Intel. Intel Open Source Computer Vision Library, v1.0. www.intel.com/research/mrl/research/opencv, October 2006.

[19] R. Jenkins, A. M. Burton, and D. White. Face recognition from unconstrained images: Progress with prototypes. In *Proceedings of the International Conference on Automatic Face and Gesture Recognition*, Southsampton, UK, April 2006.

[20] R. Kemp, N. Towell, and P. G. When seeing should not be believing: Photographs, credit cards and fraud. *Applied Cognitive Psychology*, 11(3):211–222, 1997.

[21] Y. Kirby and L. Sirovich. Application of the Karhunen-Loéve procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(1), July 1990.

[22] H. Kruppa, M. Castrillón Santana, and B. Schiele. Fast and robust face finding via local context. In *Joint IEEE Internacional Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 157–164, October 2003.

[23] H. Leder, G. Schwarzer, and S. Langton. *Development of Face Processing in Early Adolescence*, chapter 5, pages 69–80. Hogrefe & Huber, 2003.

[24] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre. Xm2vtsdb: The extended m2vts database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, March 1999.

[25] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[26] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition 2005*, 2005.

[27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. TR 6264, NISTIR, January 1999.

[28] G. Pike, R. Kemp, and N. Brace. The psychology of human face recognition. In *IEEE Colloquium on Visual Biometrics*, 2000.

[29] A. Reimondo. Opencv swiki. alereimondo.no-ip.org/OpenCV/, 2007.

[30] G. Schwarzer, N. Zauner, and M. Korell. *Face Processing during the first Decade of Life*, chapter 4, pages 55–68. Hogrefe & Huber, 2003.

[31] D. Y. Tsao and W. A. Freiwald. What's so special about the average face? *Trends in Cognitive Science*, 10(9):391–393, 2006.

[32] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, 1991.

[33] V. Vapnik. *The nature of statistical learning theory*. Springer, New York, 1995.

[34] P. Viola and M. J. Jones. Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition*, pages 511–518, 2001.

[35] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):151–173, May 2004.

[36] G. Wallis and H. Bülthoff. Learning to recognize objects. *Trends in Cognitive Sciences*, 3(1):22–31, January 2000.

[37] D. Wolpert. The existence of a priori distinctions between learning algorithms. *Neural Computation*, 8(7):13911420, 1996.