FACE ANALYSIS FOR HUMAN COMPUTER INTERACTION APPLICATIONS

Javier Ruiz-del-Solar, Rodrigo Verschae, Paul Vallejos and Mauricio Correa
Department of Electrical Engineering, Universidad de Chile, Santiago, Chile

Keywords: Human-Computer interaction through face analysis, boosting, nested cascade classifiers, face detection, multiple target tracking.

Abstract: A face analysis system is presented and employed in the construction of human-computer interfaces. This system is based on three modules (detection, tracking and classification) which are integrated and used to detect, track and classify faces in dynamic environments. A face detector, an eye detector and face classifier are built using a unified learning framework. The most interesting aspect of this learning framework is the possibility of building accurate and robust classification/detection systems that have a high processing speed. The tracking system is based on extended Kalman filters, and when used together with the face detector, high detection rates with a very low false positive rate are obtained. The classification module is used to classify the faces' gender. The three modules are evaluated on standard databases and, compared to state of the art systems, better or competitive results are obtained. The whole system is and the system is implemented in AIBO robots.

1 INTRODUCTION

Face analysis plays an important role for building human-computer interfaces that allow humans to interact with computational systems in a natural way. Face information is by far, the most used visual cue employed by humans. There is evidence of specialized processing units for face analysis in our visual system. Faces allow us the localization and identification of other humans, and the interaction and visual communication with them. Therefore, if we want that humans can interact with machines with the same efficiency, diversity and complexity used in the human-human interaction, then face analysis should be extensively employed in the construction of human-computer interfaces.

Currently, computational face analysis (face recognition, face detection, eyes detection, face tracking, facial expression detection, etc.) is a very lively and expanding research field. The increasing interest in this field is mainly driven by applications related with surveillance and security. Among many other applications we can mention video conferencing, human-robot interaction, surveillance, computer interfaces, video summarizing, image and video indexing and retrieval, biometry, and drivers monitoring.

Face detection is a key step in almost any computational task related with the analysis of faces in digital images. Moreover, in many different situations face detection is the only way to detect persons in a given scene. Knowing if there is a person present on the image (or video) is an important clue about the content of the image.

In the case of human computer interaction applications, clues about the gender, age, race, emotional state or identity of the persons give important context information. When having this kind of information, the application can be designed to respond in a different way depending on who the user is. For example, it can respond according to the mood, gender or age of the user. Face recognition systems can be improved by using other clues about the face or by having specific models (for each gender or rage). Obviously for this we require, first to be able to detect the faces and to implement accurate age, gender or race classification systems.

In this general context, the aim of this paper is to propose a face analysis system, which can be used in the construction of human-computer interaction applications. The proposed face analysis system can deal (detect, track and classify) faces in dynamic

Ruiz-del-Solar J., Verschae R., Vallejos P. and Correa M. (2007).
FACE ANALYSIS FOR HUMAN COMPUTER INTERACTION APPLICATIONS.
In *Proceedings of the Second International Conference on Computer Vision Theory and Applications*, pages 23-30
DOI: 10.5220/0002069400230030

environments. It has been implemented on AIBO robots and it performs with high accuracy as it will be shown when evaluated on standard databases.

An essential requirement of this kind of system is that it must be based in a highly robust and fast face detector. Our face detector, eye detector and gender classifier are built using a unified learning framework based on nested cascades of boosted classifiers (Verschae et al. 2006b; Verschae et al. 2006a). Key concepts used in the learning framework are boosting (Schapire and Singer, 1999), nested cascade classifiers (Wu et al., 2004), and bootstrap training (Sung and Poggio, 1998). The tracking is implemented using extended Kalman filters.

The article is structured as follows. In section 2 the learning framework that is used to train the cascade classifiers is presented. In section 3 the face detector is presented and some results of its performance are outlined. In section 4 the tracking system is described and evaluated. In section 5 the implementation of the face analysis system on Aibo robots is presented. Finally, some conclusions and projections of this work are given in section 6.

2 LEARNING FRAMEWORK

Key concepts used in the learning framework are boosting (Schapire and Singer, 1999), nested cascade classifiers (Wu et al., 2004), and bootstrap training (Sung and Poggio, 1998). A detailed description of this framework is given in (Verschae et al., 2006b).

Boosting is employed for finding (i) highly accurate hypotheses (classification rules) by combining several weak hypotheses (classifiers), each one having a moderate accuracy, and (ii) self-rated confidence values that estimate the reliability of each prediction (classification).

Cascade classification uses several layers (stages) of classifiers of increasing complexity (each layer discards non-object patterns) for obtaining an optimal system in terms of classification accuracy and processing speed (Viola and Jones, 2001). This is possible because of two reasons: (i) there is an important difference in the a priori probability of occurrence of the classes, i.e. there are much more non-object than object patterns, and (ii) most of the non-objects patterns are quite different from the object patterns, therefore they can be easily discarded by the different layers. Nested cascade classification allows to obtain higher classification accuracy by the integration of the different cascade layers (Wu et al., 2004).

Other aspects employed in the proposed framework for obtaining high-performance classification systems are: using the bootstrap procedure (Sung and Poggio, 1998) to correctly define the classification boundary, LUTs (Look-Up Tables) for a fast evaluation of the weak classifiers, simple rectangular Haar-like features that can be evaluated very fast using the integral image (Viola and Jones, 2001), and LBP features (Fröba and Ernst, 2004) that are invariant against changing illumination.

2.1 Boosted Nested Cascade

A nested cascade of boosted classifiers is composed by several integrated (nested) layers, each one containing a boosted classifier. The whole cascade works as a single classifier that integrates the classifiers of every layer. A nested cascade, composed of *M* layers, is defined as the union of *M* boosted classifiers H_C^k each one defined by:

$$H_C^k(x) = H_C^{k-1}(x) + \sum_{t=1}^{T_k} h_t^k(x) - b_k \qquad (1)$$

with $H_C^k(x) = 0$ and h_t^k the weak classifiers, T_k the number of weak classifiers in layer k, and b_k a threshold value. It should be noted that a given classifier corresponds to the nesting (combination) of the previous classifiers. The output of H_C^k is a real value that corresponds to the confidence of the classifier and its computation makes use of the already evaluated confidence value of the previous layer of the cascade (see figure 1).

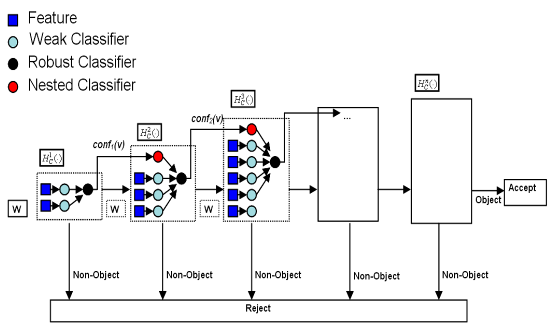

Figure 1: Block diagram of the boosted nested cascade classifier.

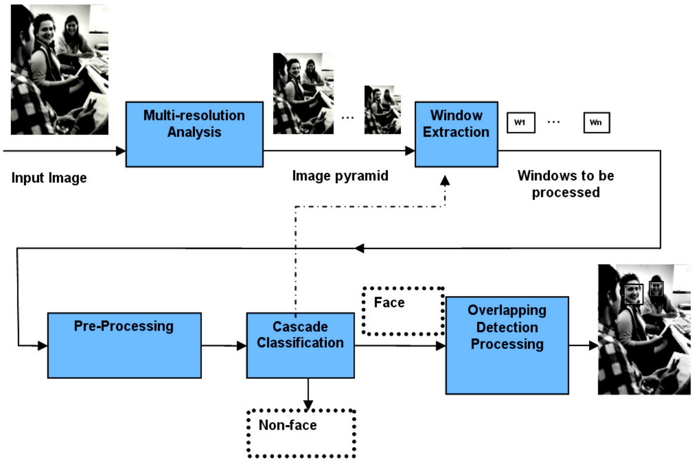

Figure 2: Block diagram of a face detection system.

Each weak classifier is applied over one feature computed in every pattern to be processed. The weak classifiers are designed after the domain-partitioning weak hypotheses paradigm (Schapire and Singer, 1999). Under this paradigm the weak classifiers make their predictions based on a partitioning of a feature domain *F*. A weak classifier *h* will have an output for each partition block, F_j, of its associated feature *f*: $h(f(x)) = c_j \ni f(x) \in F_j$. Thus, the weak classifiers prediction depends only on which block F_j a given sample (instance) falls into. For each classifier, the value associated to each partition block (c_j), i.e. its output, is calculated for minimizing a bound of the training error and at the same time a bound on an exponential loss function of the margin (Schapire and Singer, 1999). This value is given by:

$$c_j = \frac{1}{2}\ln\left(\frac{W_{+1}^j + \varepsilon}{W_{-1}^j + \varepsilon}\right) \quad (2)$$

$$W_l^j = \Pr[f(x) \in F_j \wedge y_i = l], l = \pm 1 \quad (3)$$

and ε a regularization parameter (Schapire and Singer, 1999).

A slightly modified version of the real Adaboost learning algorithm (Verschae et al. 2006b) is employed for selecting the features and training the weak classifiers taking into account the nested configuration of the cascade.

3 DETECTION SYSTEM

In the following we briefly present the developed face detection system. The block diagram of the face detection systems is presented in figure 2. First, for detecting faces at different scales a multiresolution analysis is performed by scaling the input image by a factor of 1.2 (*Multiresolution Analysis* module). This scaling is performed until images of about 24x24 pixels are obtained. Afterwards, windows of 24x24 pixels are extracted in the *Window Extraction* module for each of the scaled versions of the input image. The extracted windows can be then pre-processed for obtaining invariance against changing illumination. Thanks to the use of features which are invariant against changing illumination to a large degree we do not perform any kind of preprocessing.

Afterwards, the windows are analyzed by the nested cascade classifier (*Cascade Classification Module*) built with the framework described in section 2. Finally, in the *Overlapping Detection Processing* module, the windows classified as faces are fused (normally a face will be detected at different scales and positions) for obtaining the size and position of the final detections. This fusion is described in (Verschae and Ruiz-del-Solar, 2003).

The eye detector works in the same was as the face detector does, the only difference is that the search is not done within the whole image, but only within the face. As the face detector, the eye detector woks on 24x24 windows, therefore it can be used only on faces of 50x50 pixels or larger.

The gender classifier was built using the learning framework as the eye and face detectors. The gender classifier works on windows of 24x24 pixels and when the eye positions are available it uses them for aligning the faces. In (Verschae et al. 2006a) we give a detailed description and evaluation of the gender classifier.

Table 1: Comparative evaluation (DR: Detection Rate) of the face detector on the BioID Database (1,521 images).

False Positives	**DB**	**0**	**1**	**2**	**3**	**6**	**15**	**17**	**25**
Out Method	UCHILE			87.8	88.0	94.8		98.5	
Out Method	FERET	98.7	99.5		99.7				
Out Method	BIOID	94.1	95.1	96.5		96.9	97.6		98.1
Fröba and Ernst 2004	BIOID		~50			~65	~84		~98

Table 2: Comparative evaluation (DR) of the face detector on the CMU-MIT database (130 images, 507 faces). Notice that in (Fröba et al. 2004) a subset of 483 (out of 507) faces is considered. This subset is called CMU 125 testset.

False Positives	**0**	**3**	**5**	**6**	**10**	**13**	**14**	**19**	**25**	**29**	**31**	**57**	**65**
Our Method		77.3	83.2				86.6	88	89.9		90.1		92.1
Fröba et al. 2004	~66		~87						~90				
Wu et al. 2004		**89**			90.1	**90.7**						**94.5**	
Viola and Jones 2001					76.1						88.4		92
Rowley et al.1998					83.2						86		
Schneiderman 2004				**89.7**				**93.1**		**94.4**			
Li. et al. 2002					83.6						90.2		
Delakis and Garcia 2004	**88.8**				**90.5**						91.5		92.3

For testing purposes we employed four databases (BIOID, 2005), FERET (Phillips et al. 1998), CMU-MIT (Rowley et al 1998), and (UCHFACE, 2006). No single image from these databases was used for the training of our systems. Selected examples of our face detection, at work in the FERET, BIOID, UCHFACE and MIT-CMU databases, are shown in figure 3. The figures also show eyes detection and gender classification.

The face detector was evaluated using two types of databases: (a) BIOID and FERET, which contain one face per image, and (b) CMU-MIT and UCHFACE, which contain none, one or more faces per image. Table 1 shows results of our method for the FERET, BIOID and UCHILE databases as well as the results for (Fröba and Ernst 2004) for the BIOID database.

In the BIOID database, which contains faces with variable expressions and cluttered backgrounds, we obtain a high accuracy, a 94.1% detection rate with zero false positives (in 1521 images), while on the FERET database, which contains faces with neutral expression and homogeneous background, we obtain a very high accuracy, a 99.5% detection rate with 1 false positive (in 1016 images). These results were obtained without considering that there is only one face per image.

In the UCHFACE database (343 images), which contains faces with variable expressions and cluttered backgrounds, we consider that the obtained results are rather good (e.g. 88.0% with 3 false positives, 98.5% with 17 false positives).

The table 2 shows comparative results with state of the art methods fot the CMU-MIT database. In the CMU-MIT database we also obtain good results (e.g. 83.2% with 5 false positives and 88% with 19 false positives). If we compare to state of the art methodologies in terms of DR and FP, we obtain better results than (Viola and Jones, 2001; Rowley et al, 1998), slightly better results than (Li et al, 2002), slightly worse results than (Delakis and Garcia, 2004) (but our system is is about 8 times faster), and worse results than (Wu et al. 2004) and (Schneiderman, 2004). We think we have lower detection rates than (Wu et al. 2004) and (Schneiderman, 2004) mainly because of the size of the training database. For example in (Wu et al. 2004) 20,000 training faces are employed while our training database consists of 5,000 face images. Notice that our classifier is among the fastest ones. The ones that have a comparable processing time are (Viola and Jones 2001), (Fröba et al. 2004), (Wu et al. 2004) and (Li. et al. 2002).

The gender classifier performance was evaluated in two cases: when the eyes were manually annotated and when the eyes were automatically detected. Table 3 shows results of this evaluation for the UCHFACE. FERET and BIOID databases. It is should be noticed that its behaviour is very robust to changes in the eyes positions that are used for the face alignment and that in two of the databases best results are obtained when the eye detector is used.

Table 3: Gender classification results: Percentage of correct classification when eyes are annotated or detected.

Database	Annotated eyes	Detected eyes
UCHFACE	81.23 %	80.12%
FERET	85.56 %	85.89%
BIOID	80.91 %	81.46%

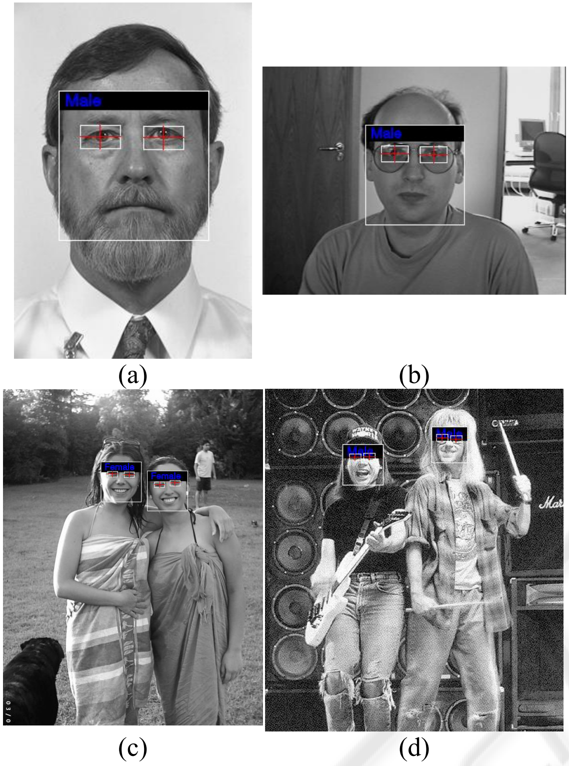

Figure 3: Selected examples of our face detection, eyes detection and gender classification systems at work on the FERET (a), BIOID (b), UCHFACE (c) and MIT-CMU (d) databases.

4 FACE TRACKING USING KALMAN FILTERS

The tracking of the faces is based mainly on the use of Extended Kalman Filters (EKFs). Although from the theoretical point of view it can be argued that Particle Filters (e.g. (Isard and Blake 1998) are superior than EKF because of the Gaussianity hypothesis (Dudek, and Jenkin, 2002), our experience with self-localization algorithms for mobile robotics (Lastra et al., 2004) tell us that the performance of both kind of filters in tracking and self-localization tasks is rather similar. Moreover, it is possible to obtain a very fast implementation of the EKF if the state vector is small, as in our case, because for each tracked object a different EKF is employed. This is very important when several objects are tracked at the same time.

4.1 State Vectors and Parameters Database

Each object (face) is characterized by its position in pixels in the frame, its width, its height, and the corresponding changing rates of these variables. The eight variables are the state vector of a first order EKF ($\mathbf{x}_k$). The parameters database (DB) stores the latest state vector ($\mathbf{x}_{k-1}$) for each object under tracking and its associated EKF. Since the detected features do not include the change rate components, these components are estimated as:

$$\mathbf{z}_k^{\mathbf{T}} = \begin{pmatrix} z_k^1 & z_k^2 & z_k^3 & z_k^4 & \frac{z_k^1 - x_{k-1}^1}{\Delta t} & \frac{z_k^2 - x_{k-1}^2}{\Delta t} & \frac{z_k^3 - x_{k-1}^3}{\Delta t} & \frac{z_k^4 - x_{k-1}^4}{\Delta t} \end{pmatrix}^{\mathbf{T}} \quad (4)$$

With $\mathbf{z}_k$ the vector of observations. The update model is:

$$\hat{\mathbf{x}}_k = \left(\mathbf{I} - \Delta t \begin{pmatrix} 0 & .. & 0 & 1 & 0 & 0 & 0 \\ & & & 0 & 1 & 0 & 0 \\ & & & & 0 & 1 & 0 \\ : & & & .. & & 0 & 1 \\ & & & & & & 0 \\ & & & & & & : \\ 0 & & & .. & & & 0 \end{pmatrix} \right) \hat{\mathbf{x}}_{k-1} \quad (5)$$

4.2 Tracking Procedure

The block diagram of the multiple face detection and tracking system is shown in figure 4. Input images are analyzed in the *Face Detector* module, and detected faces are further processed by the *Detected-Tracked Object Matching* module. In this module the detected faces are matched with the current objects under tracking. Each new detection (a face window) is evaluated in the Gaussian function described by the state vector and its covariance matrix on the Kalman filter of each object. In this way, a matching probability is calculated. If the matching probability is over a certain threshold, the detected face is associated with the corresponding object. If no object produces a probability value over that threshold, then the detected face is a new candidate object, and a new state vector (and Kalman filter) is created for this new object (New Object Generator module).

Table 4: Face detection results on the sets A and B from PETS-ICVS 2003.

Set A	Detection Rate [%]	67,9	62,1	50,8	44,9	36,2
	# False Positives	851	465	334	292	242
Set B	Detection Rate [%]	67,2	60,9	53,5	44,7	37,9
	# False Positives	88	50	37	32	22

Table 5: Face detection results, after tracking, on set A from PETS-ICVS 2003.

Tracking Parameters			False Positive	Detection Rate [%]	False Positive Decrement	Detection Rate Increment
MCF	MCF					
3	5	2.0	525	65,19	35,0%	10,7%
3	7	2.0	530	65,23	34,4%	11,0%
2	2	2.0	536	65,71	33,7%	11,8%
2	5	2.0	580	66,76	28,2%	13,6%
2	7	2.0	580	66,76	28,2%	13,6%
3	5	1.0	625	68,51	22,6%	16,5%
3	7	1.0	655	68,96	18,9%	17,3%
2	2	1.0	629	68,74	22,2%	16,9%
1	2	2.0	683	68,31	15,5%	16,2%
1	5	2.0	700	68,59	13,4%	16,7%
1	7	2.0	700	68,59	13,4%	16,7%
2	5	1.0	738	70,10	8,7%	19,2%
2	7	1.0	750	70,23	7,2%	19,5%

Table 6: Face detection results, after tracking, on set B from PETS-ICVS 2003.

Tracking Parameters			False Positive	Detection Rate[%]	False Positive Decrement	Detection Rate Increment
CTO	MCF	Q				
3	5	1.0	69	68,9	21,6%	2,5%
3	7	1.0	71	68,9	19,3%	2,5%
2	2	1.0	71	689	19,3%	2,5%
1	2	2.0	72	68,0	18,2%	1,2%
1	5	2.0	76	68,1	13,6%	1,3%
1	7	2.0	76	68,1	13,6%	1,3%
2	5	1.0	80	69,8	9,1%	3,9%
2	7	1.0	80	69,9	9,1%	4,0%
3	5	0.5	85	70,7	3,4%	5,2%
2	2	0.5	87	70,6	1,1%	5,1%
3	7	0.5	88	70,7	0,0%	5,2%

For each object under tracking, the prediction model estimates its a priori state (Object State Prediction module). Then, the a priori state is updated using all the detections associated with this state in the matching stage (Object Update module).

If any candidate object accomplish the promote rule (over a certain amount of detections in a maximal amount of frames) then it becomes a true object (Candidate Promoter module). Finally, if a candidate object has more than a certain amount of frames with not enough associated detections (below a certain threshold), it is eliminated from the database (*Object Filter* module). True objects with state probability below a certain threshold are also eliminated from the database.

4.3 Multiple Detection and Tracking in Dynamic Environments

We have integrated the face detection and tracking system, building a system for the tracking of multiple faces in dynamic environments. As it will be shown, this system is able to detect and track faces with a high performance in real-world videos, and with an extremely low number of false positives compared to state of the art methodologies.

In order to test performance of our multiple face detection and tracking system we employed the PETS-ICVS 2003 dataset. The PETS initiative corresponds to a very successful series of workshops on Performance Evaluation of Tracking and Surveillance. The PETS 2003 topic was gesture and action recognition, more specifically the annotation of a "smart meeting" (includes facial expressions, gaze and gesture/action). The PETS-ICVS 2003 dataset (PETS, 2003) consists of video sequences (frame from 640x480 pixels) captured by three cameras on a conference room. Two cameras (camera 1 and 2) were placed on opposite walls capturing the participant on each side of the room, and the third camera (camera 3) is an omnidirectional camera on the desk center. The dataset is divided in four scenarios A, B, C and D. For this analysis frames from scenarios A, B, and D, and cameras 1 and 2, were used. The ground truth consists of the eyes coordinates for those frames divisible by 10. In our experiments all frames were processed, but for statistics just two sets of images were considered: (i) Set A: all annotated frames, i.e. frames with frame number divisible by 10, and (ii) Set B: frames with frame number divisible by 100. The set A contains 49,350 frames and 10,308 annotated faces, while the set B contains 4,950 frames and 1,000 annotated faces.

In table 4 are shown the detection results obtained by our face detector (without the tracking) on both sets. These results are much better than the ones reported in (Cristinacce and Cootes, 2003), where the Fröba-Kullbeck detector (Fröba and Küblbeck, 2002) and the Viola&Jones detector (Viola and Jones, 2001) were tested on the set B. In that test the Viola&Jones detector outperforms the Fröba-Kullbeck, but the results it obtains are very poor, 50% DR with 202 false positives or 62.2% DR with 2,287 false positives. We can conclude that our face detector performs very well in this real-world dataset (4,950 frames), and that the amount of FP is extremely low.

We analyzed the performance of our tracking system and we quantify the improvement in the face detection process when using such a system. We have analyzed the behavior of three different parameters of the tracking system:

- CTO (Candidate to true object) threshold: A new detection is immediately added to the database in order to track it, but it is not considered as a true tracked object until CTO other detections are associated with it.
- MCF (Max candidate frames): If a candidate to object does not reach the CTO threshold in MCF frames since it was added to the database, it is eliminated.
- Q: This is the covariance of the process noise in the Kalman filter.

In table 5 and 6 are shown the detection results after applying the tracking system on sets A and B. It can be seen that thanks to the tracking, the number of FP decreases largely, up to 21% in set B and 35% in set A, and that at the same time the DR increases.

5 PERSON DETECTION AND TRACKING FOR AIBO ROBOTS

Sony AIBO robots correspond to one of most widespread and popular personal robots. Thousands of children and researchers employ AIBO robots for entertainment or research. We believe that in a near future personal robots will be far more widespread than today. One of the basic skills that personal robots should integrate is the face-based visual interaction with humans. Robust face analysis is a key step in this direction. For implementing such a system we adapted the face analysis system already described for Sony AIBO robots model ERS7.

ERS7 robots have a 64bit RISC Processor (MIPS R7000) from 576 MHz, 64MB RAM and a color camera of 416x320 pixels that delivers 30fps. The face and tracking detection system was integrated with our robot control libraryU-Chile1 (Lastra et al 2004)(Ruiz-del-Solar et al, 2005b). U-Chile1 is divided in five task-oriented modules: Vision, which contains mostly low-level vision algorithms, Localization, in charge of the robot self-localization, Low-level Strategy, in charge of the behavior-based control of the robots, High-level Strategy, in charge of the high level robot behavior and strategy, and Motion Control, in charge of the control of the robot movements. U-Chile1 runs in real-time and after the integration with the tracking system we were able of running our face detection and tracking system at a rate of 2fps. We included also eyes detection and gender classification. In figure 5 are shown some selected examples of face detection and tracking using an AIBO ERS7. The system detects faces, gender classification and eyes detection. More examples can be seen in (UCHFACE, 2005).

6 CONCLUSIONS

Face analysis plays an important role for building human-computer interfaces that allow humans to interact with computational systems in a natural way. Face information is by far, the most used visual cue employed by humans. In this context in the present article we have proposed face analysis system that can be used to detect, track and classify (gender) faces. The proposed system can be used in the construction of different human-computer interfaces.

The system is based on a face detector with high accuracy, a high detection rate with a low number of false positives. This face detector obtains the best-reported results in the BioID database, the best reported results in the PETS-ICVS 2003 dataset, and the third best reported results in the CMU-MIT database.

The face detector was integrated with a tracking system for building a system for the tracking of multiple faces in dynamic environments. This system is able to detect and track faces with a high performance in real-world videos, and with an extremely low number of false positives compared to state of the art methodologies. We also integrated our face analysis and tracking system into the Sony AIBO robots. In this way the AIBO robots can interact with persons using the human faces and the gender and eye information.

Beside the already mentioned projects, we are currently applying our face analysis system for developing a service robot that interacts with humans using face information and on a retrieval tool for searching persons on image and video databases.

ACKNOWLEDGEMENTS

This research was funded by Millenium Nucleus Center for Web Research, Grant P04-067-F, Chile. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

REFERENCES

BioID, 2005.Face Database. Available on August 2005 in: http://www.bioid.com/downloads/facedb/index.php

Cristinacce, D. and Cootes, T., 2003. “A Comparison of two Real-Time Face Detection Methods”, 4th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance – PETS 2003, pp. 1-8, Graz, Austria.

Fröba, B. and Küblbeck, Ch., 2002. Robust face detection at video frame rate based on edge orientation features. 5th Int. Conf. on Automatic Face and Gesture Recognition FG - 2002, pp. 342–347.

Fröba, B. and Ernst, A., 2004. “Face detection with the modified census transform”, 6th Int. Conf. on Face and Gesture Recognition - FG 2004, pp. 91–96, Korea.

Delakis, M. and Garcia, C., 2004. “Convolutional face finder: A neural architecture for fast and robust face detection”, IEEE Trans. Pattern Anal. Mach. Intell., 26(11):1408 – 1423.

Dudek, G., and Jenkin, M., 2002. Computational Principles of Mobile Robotics, Cambridge University Press.

Isard, M., and Blake, A., 1998. “Condensation – Conditional Density Propagation for Visual Tracking”, Int. J. of Computer Vision, Vol. 29, N. 1, pp. 5—28.

Lastra, R., Vallejos, P., and J. Ruiz-del-Solar, 2004. “Integrated Self-Localization and Ball Tracking in the Four-Legged Robot Soccer League”, 1st IEEE Latin American Robotics Symposium – LARS 2004, pp. 54 – 59, Oct. 28 – 29, México City, México.

Li, S.Z., Zhu, L., Zhang, Z.Q., Blake, A., Zhang, H.J. and Shum, H., 2002. “Statistical Learning of Multi-view Face Detection”, 7th European Conf. on Computer Vision – ECCV 2002, (Lecture Notes in Computer Science; Vol. 2353), pp. 67 – 81.

PETS, 2003. Home Page. Available on August 2005 in: http://www.cvg.cs.rdg.ac.uk/PETS-ICVS/pets-icvs-db.html

Phillips, P. J., Wechsler, H., Huang, J.and Rauss, P., 1998. “The FERET database and evaluation procedure for face recognition algorithms,” Image and Vision Computing J., Vol. 16, no. 5, pp. 295-306.

Rowley, H., Baluja, S. and Kanade, T., 1998. “Neural Network-Based Detection”, IEEE Trans. Pattern Anal. Mach. Intell., Vol.20, No. 1, 23-28.

Ruiz-del-Solar, J., Vallejos, P., Lastra, R., Loncomilla, P., Zagal, J.C., Morán, C., and Sarmiento, I., 2005b. “UChile1 2005 Team Description Paper”, RoboCup 2005 Symposium, July 13 – 17, Osaka, Japan.

Schapire, R.E. and Singer, Y., 1999. Improved Boosting Algorithms using Confidence-rated Predictions, Machine Learning, 37(3):297-336.

Schneiderman, H., 2004. Feature-Centric Evaluation for Efficient Cascade Object Detection, IEEE Conference of Computer Vision and Pattern Recognition – CVPR 2004, pp. 29 - 36.

Sung, K. and Poggio, T., 1998. “Example-Based Learning for Viewed-Based Human Face Deteccion”, IEEE Trans. Pattern Anal. Mach. Intell., Vol.20, N. 1, 39-51.

UCHFACE, 2005. Computational Vision Group, Universidad de Chile. Available on August 2005 in: http://vision.die.uchile.cl/

Verschae, R. and Ruiz-del-Solar, J., 2003. “A Hybrid Face Detector based on an Asymmetrical Adaboost Cascade Detector and a Wavelet-Bayesian-Detector”, Lecture Notes in Computer Science 2686 (IWANN 2003), Springer, pp. 742-749, Menorca, Spain.

Verschae, R., Ruiz-del-Solar, J., Correa, M., 2006a. Gender Classification of Faces Using Adaboost, In Proc of CIARP, LNCS 4225.

Verschae, R., Ruiz-del-Solar, J., Correa, M. and Vallejos, P. 2006b. A Unified Learning Framework for Face, Eyes and Gender Detection using Nested Cascades of Boosted Classifiers, Tech.Report UCH-DIE-VISION-2006-02, Dept. of Elect. Eng., Universidad de Chile.

Viola, P. and Jones, M., 2001. "Rapid object detection using a boosted cascade of simple features", IEEE Conf. on Computer Vision and Pattern Recognition - CVPR, Kauai, HI, USA, 2001, pp. 511 - 518.

Wu, B., AI, H. , Huang, C. and Lao, S., 2004. “Fast rotation invariant multi-view face detection based on real Adaboost”, 6th Int. Conf. on Face and Gesture Recognition - FG 2004, pp. 79–84, Seoul, Korea.

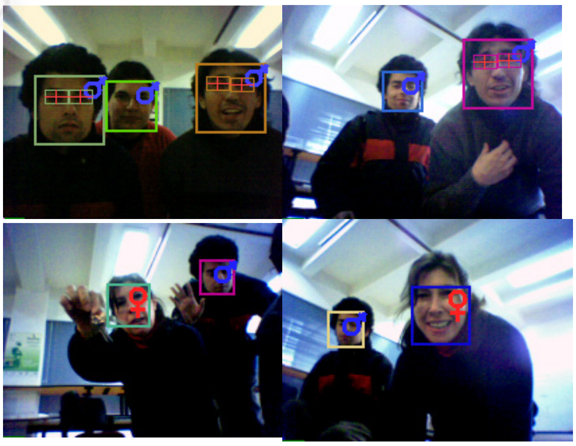

Figure 5: Examples of the face detection and tracking system for AIBO robots. The system detect faces and performs gender classification. When the resolution of the faces is larger than 50x50 pixels it detects also the eyes.