

FACE AND FACIAL FEATURE DETECTION EVALUATION

Performance Evaluation of Public Domain Haar Detectors for Face and Facial Feature Detection

M. Castrillón-Santana, O. Déniz-Suárez, L. Antón-Canalís and J. Lorenzo-Navarro
Institute of Intelligent Systems and Numerical Applications in Engineering
Edificio Central del Parque Tecnológico, Campus de Tafira, University of Las Palmas de Gran Canaria, Spain

Keywords: Face and facial feature detection, haar wavelets, human computer interaction.

Abstract: Fast and reliable face and facial feature detection are required abilities for any Human Computer Interaction approach based on Computer Vision. Since the publication of the Viola-Jones object detection framework and the more recent open source implementation, an increasing number of applications have appeared, particularly in the context of facial processing. In this respect, the OpenCV community shares a collection of public domain classifiers for this scenario. However, as far as we know these classifiers have never been evaluated and/or compared. In this paper we analyze the individual performance of all those public classifiers getting the best performance for each target. These results are valid to define a baseline for future approaches. Additionally we propose a simple hierarchical combination of those classifiers to increase the facial feature detection rate while reducing the face false detection rate.

1 INTRODUCTION

Fast and reliable face and facial element detection are topics of great interest to get more natural and comfortable Human Computer Interaction (HCI) (Pentland, 2000). Therefore, the number of approaches addressing this problem have increased in the last years (Li et al., 2002; Schneiderman and Kanade, 2000; Viola and Jones, 2004) providing reliable approaches to the Computer Vision community. However, since the recent work by Viola and Jones (Viola and Jones, 2004) describing a fast multi-stage general object classification approach, and the release of an open source implementation (Lienhart and Maydt, 2002), this approach has been extensively used in Computer Vision research, particularly for detecting faces and their elements. Different authors have made their classifiers (not their training sets) public. However, as far as the authors of this paper know, no performance evaluation has been done yet.

In this paper, we compare different public domain classifiers, based on Lienhart's implementation, related to face and facial element detection, in order to provide a baseline for future developments. Section 2 briefly introduces the detection approach. Sections 3

and 4 present respectively the experimental setup and the conclusions extracted.

2 HAAR-BASED DETECTORS

The general object detector framework described in (Viola and Jones, 2004), is based on the idea of a boosted cascade of weak classifiers. For each stage in the cascade, a separate subclassifier is trained to detect almost all target objects while rejecting a certain fraction of the non-object patterns (which were accepted by previous stages).

The resulting detection rate, D , and the false positive rate, F , of the cascade are given by the combination of each single stage classifier rates:

$$D = \prod_{i=1}^K d_i \quad F = \prod_{i=1}^K f_i \quad (1)$$

Each stage classifier is selected considering a combination of features which are computed on the integral image. These features are reminiscent of Haar wavelets and early features of the human visual

pathway such as center-surround and directional responses. The implementation (Lienhart and Maydt, 2002) integrated in the OpenCV (Intel, 2006) extends the original feature set (Viola and Jones, 2004).

With this approach, given a 20 stage detector designed for refusing at each stage 50% of the non-object patterns (target false positive rate) while falsely eliminating only 0.1% of the object patterns (target detection rate), its expected overall detection rate is $0.999^{20} \approx 0.98$ with a false positive rate of $0.5^{20} \approx 0.9 \cdot 10^{-6}$. This schema allows a high image processing rate, due to the fact that background regions of the image are quickly discarded, while spending more time on promising object-like regions. Thus, the detector designer chooses the desired number of stages, the target false positive rate and the target detection rate per stage, achieving a trade-off between accuracy and speed for the resulting classifier.

3 EXPERIMENTS

3.1 Experimental Setup

The available collection of public domain classifiers that we have been able to locate are described in Table 1. Their targets are frontal faces, profile faces, head and shoulders, eyes, noses and mouths. Reference information is included as well as the size of the smallest detectable pattern, the label used below in the figures, and the processing time in seconds needed to analyze the whole test set.

In order to analyze the performance of the different classifiers, the CMU dataset (Schneiderman and Kanade, 2000) has been chosen for the experimental setup. This dataset contains a collection of heterogeneous images, feature that from our point of view provides a better understanding of the classifier performance. The dataset is divided into four different subsets *test*, *newtest*, *test-low* and *rotated* combining the test sets of Sung and Poggio (Sung and Poggio, 1998) and Rowley, Baluja, and Kanade (Rowley et al., 1998). The dataset and the annotation data can be obtained at (Carnegie Mellon University, 1999).

3.2 Face Detection

In this Section the detection performance of those classifiers designed to detect the whole face or head is presented. We have considered the different frontal face detection classifiers provided with the OpenCV release (Lienhart et al., 2003a), the profile face detector (Bradley, 2003), and the head and shoulders detector (Kruppa et al., 2003).

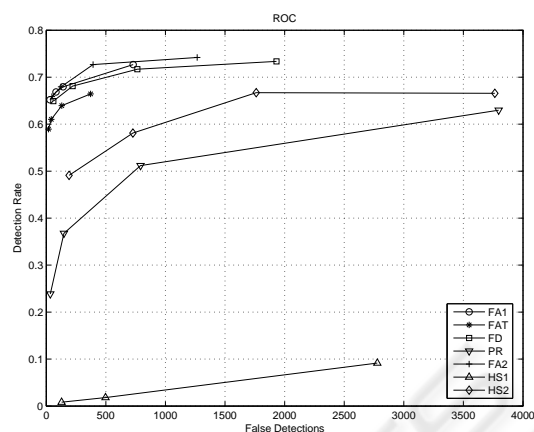


Figure 1: Face and head and shoulders detection performance.

The criterion to determine if a face is correctly detected requires that all its features are located inside the detection, and the detection container width must not be greater than four times the distance between the annotated eyes.

For each classifier its ROC curve was computed applying the original release and some variations obtained reducing its number of stages. Observing the Area Under the Curve (AUC) of the resulting ROC curves, see Figure 1, *FA2* (*haarcascade_frontalface_alt2* in the OpenCV distribution (Intel, 2006)) offered the best performance closely followed by *FA1* and *FD* (respectively *haarcascade_frontalface_alt* and *haarcascade_frontalface_default*). It is, additionally, the fastest among them. Remember that the test dataset used is composed by four different subsets. The original *FA2* classifier provided for the different subsets the performance showed in the corresponding column of Table 2. Note the lower performance achieved for the rotated subset. These results were expected as the Viola-Jones' framework is not able to accept more than slight variations in rotation with respect to the positive samples used.

Table 2: Frontal face and profile detection performance for each subset.

Subset	Frontal face detection performance	Profile detection performance
newtest	89.07%	62.30%
test	86.98%	43.20%
rotated	19.28%	10.76%
test-low	83.56%	36.99%

The single profile detector available reported lower and less homogeneous performance, showing

Table 1: Public domain classifiers available.

Target	Authors/Reference	Download	Size	Stages	Proc. time (secs.)	Label
Frontal faces	(Lienhart et al., 2003b; Lienhart et al., 2003a)	(Intel, 2006)	24 × 24	25	37.6	FD
Frontal faces	(Lienhart et al., 2003b; Lienhart et al., 2003a)	(Intel, 2006)	20 × 20	21	41.8	FA1
Frontal faces	(Lienhart et al., 2003b; Lienhart et al., 2003a)	(Intel, 2006)	20 × 20	46	31.8	FAT
Frontal faces	(Lienhart et al., 2003b; Lienhart et al., 2003a)	(Intel, 2006)	20 × 20	20	35.7	FA2
Profile faces	(Bradley, 2003)	(Reimondo, 2007)	20 × 20	26	44.8	PR
Head and shoulders	(Kruppa et al., 2003)	(Intel, 2006)	22 × 18	30	66.5	HS1
Head and shoulders	(Kruppa et al., 2003)	TBP	22 × 20	19	98.4	HS2
Left eye	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	18 × 12	16	44.9	LE
Right eye	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	18 × 12	18	44.8	RE
Eye	M. Wimmer	(Wimmer, 2004)	25 × 15	5	12.1	E1
Eye	Urtho	(Urtho, 2006)	10 × 6	20	22	E2
Eye	Ting Shan	(Reimondo, 2007)	24 × 12	104	20.9	E3
Eye pair	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	45 × 11	19	14.6	EP1
Eye pair	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	22 × 5	17	15.4	EP2
Eye pair	(Bediz and Akar, 2005)	(Reimondo, 2007)	35 × 16	19	16.7	EP3
Nose	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	25 × 15	17	27	N1
Mouth	(Castrillón Santana et al., 2007)	(Reimondo, 2007)	25 × 15	19	41.4	M1
Mouth	(Liang et al., 2002)	(Intel, 2006)	32 × 18	18	15.6	M2

a better performance for the first subset as suggests the last column of Table 2 for the original classifier.

Paying attention to the head and shoulders detector, *HS1*, in Figure 1, its results evidence a rather low performance, as seen in the Figure. For that reason we used a more recently trained classifier (*HS2*) following the same approach but with a larger training dataset. Its performance was notoriously better, circumstance that made us to suspect the presence of a bug in the classifier included in the current OpenCV release. It must be noticed that this classifier requires the presence of the local context to react, circumstance that does not occur for different images contained in the dataset. It must be also observed that both classifiers provided a higher false detection rate than those trained to detect frontal faces.

3.3 Facial Element Detection

A similar experimental setup was carried out for the classifiers specialized in facial feature detection such as eyes, nose and mouth. Figures 2 shows the results achieved using the different eye detectors for the left eye (the results obtained for the right eye are quite similar, and therefore not included here).

The best performance is given by *RE* and *LE*, but requiring more processing time. Both classifiers provide similar performance, providing always a lower false detection rate (except for *E2*) and a greater detection rate. For all the classifiers, we have not considered the detection of the opposite eye as a false detection. For *LE* and *RE* the classifiers performed worse detecting the other eye, but they always got greater detection rate than any other classifier.

A facial feature is considered correctly detected, if the distance to the annotated location is less than 1/4 of the actual distance between the annotated eyes. This criterion was used to estimate the eye detection success originally in (Jesorsky et al., 2001).

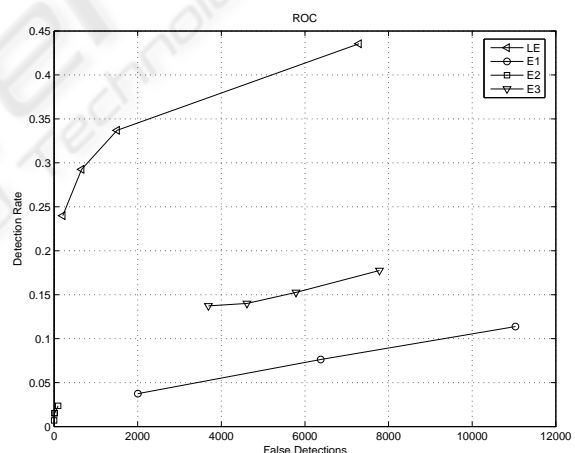


Figure 2: Left eye detection performance.

The performance achieved by *LE* for the different subsets is showed in Table 3. It is observed that this feature is also detected, with a similar rate, when faces are rotated, showing even a higher detection rate than for frontal faces. However, facial feature detectors perform poorly with low resolution (*test-low* set), i.e. they require larger faces to be reliable.

We have also analyzed the eye pair, nose and mouth detectors, see Figure 3. The best eye pair detection performance is given by *EP3*, with similar processing cost. The nose detector has the lowest performance among the whole set of facial feature de-

Table 3: Left eye detection for each subset.

Subset	Detection performance
newtest	39.89%
test	29.69%
rotated	32.29%
test-low	10.96%

tectors analyzed. However the best mouth detector, behaves notoriously better.

As a summary, the best facial element detectors available in the public domain are: *RE*, *LE*, *EP3*, *N* (only one available) and *MI*.

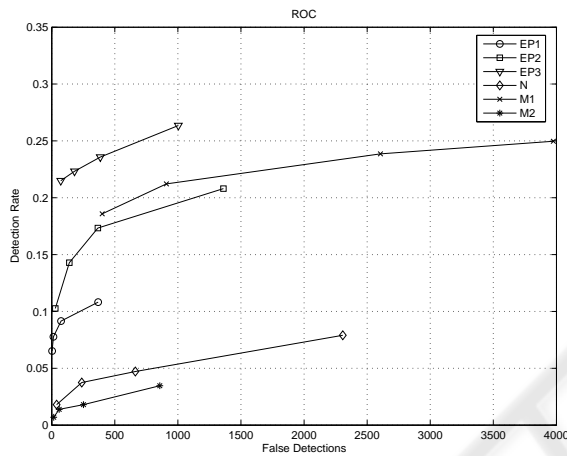


Figure 3: Eye pair (*EP1-3*, Nose (*N*) and Mouth (*M1* and *M2*) detection performance.

Even when searching facial features in the whole image reported much lower performance than for face detectors, it is evidenced that in some circumstances face detection could be tackled by detecting some of its inner features. An example of this is shown in Figure 4. The best face detector used was not able to locate the evident face, likely because it is partially cropped. However some of their inner features were positively detected. Certainly, some false detections appear, but geometric consistency is a simple approach to filter them (Wilson and Fernandez, 2006). The main drawback is that the processing time will be larger.

3.4 Combined Detection

As shown in previous sections, facial element detection seems to behave worse than facial detection following the same searching approach. A reasonable reason for this is the lower resolution of facial elements in relation to the face, therefore the different classifiers are trying to locate patterns which are smaller than those used for training. In this section we

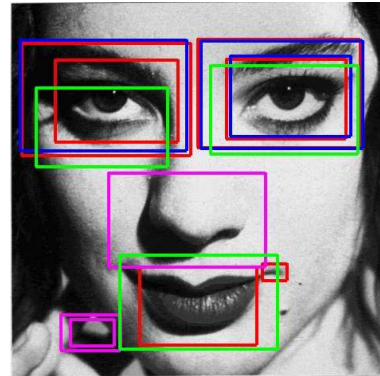


Figure 4: Sample image of the CMU dataset (?) that reported no face detection. However, facial feature detection provided a set of detections (red and blue for left and right eye respectively, magenta for nose and green for mouth). Even when some false detection are present, they could be in many situations filtered by means of the coherent geometric location of the different features (Wilson and Fernandez, 2006).

will apply facial feature detection only in those areas where a face has been detected, in order to analyze if their performance can be improved, and additionally, to observe if the no detection of facial elements can be used as a filter to remove some false face detections.

To detect faces we used the *FA2* detector with 18 stages, see Table 1, which in Section 3.2 evidenced the best relation between processing speed and detection ratio. Once a face is given, the individual facial elements were searched using three different procedures: 1) Searching the facial elements in those areas or regions of interest (ROIs), where a face has been detected (C1), 2) Scaling up the ROI image twice (in order to have bigger facial details) before searching (C2), and 3) Scaling up the ROI twice and bounding the acceptable detections size according to the detected face size (C3). We would like to remark that for every approach each feature is searched always in a ROI coherent with the expected feature location for a frontal face, similarly to (Wilson and Fernandez, 2006).

Figure 5 compares the results achieved for the *LE* classifier using the different approaches: C1, C2 and C3. With C1, we just restricted the search area to the face detected, the facial feature detection rate is reduced in relation to the results presented in Section 3.3.

Scaling up the ROIs, C2 and C3, increases significantly the facial feature detection rate, but also the false detection rate, as suggested by Figure 5. However, the latter keeps being lower than in absence of ROIs, see Figure 2. In addition, if we relate to the face size the acceptable size of the facial features to search, C3, the false detection rate is clearly reduced

keeping a similar detection rate. Due to the reduced space available, the results for the other facial features are summarized in Table 4.

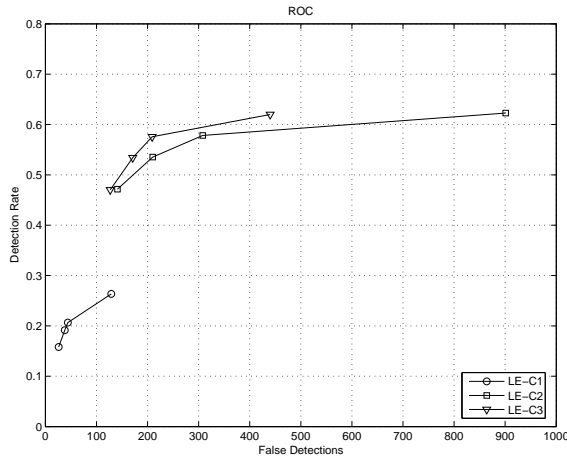


Figure 5: LE performance for the different approaches.

Table 4: Facial feature detection performance with different approaches (TD: correct detection rate, FD: number of false detections).

App.	Left eye		Right eye		Nose		Mouth	
	TD	FD	TD	FD	TD	FD	TD	FD
C1	17.2%	18	15.1%	18	1.7%	0	17.1%	52
C1	49.1%	154	45.5%	137	22.2%	120	41.5%	196
C3	48.7%	128	45.1%	124	21.9%	104	41.5%	184

We have also applied a soft heuristic to reduce the number of false face detections, making use of the results achieved for the facial features. Big enough face detections, larger than 40 pixels in width, that report no facial feature detection are not accepted as face detections. Figure 6 presents the results achieved for FA2 and different combinations of the facial features detectors. The original performance is depicted with FA, the modified with FAv. The detection rate is almost not affected, while achieving an improvement in false detection reduction. To give an example, for the most restrictive FA2 classifier, the initial number of false detections was 47, but applying this approach was reduced to 35.

Are there false detections with more than one facial element detected? The answer is positive for four images. An example of is presented in Figure 7, only the human faces are annotated, but Peggy is there. These detections were labeled as incorrect only because the container is a little bit bigger than it should. Indeed the facial features could be used to better fit the container.

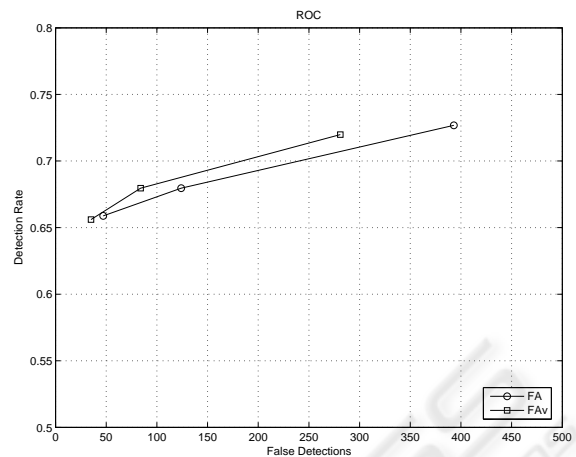


Figure 6: Face detection results achieved with approaches FA and FAv. Note the reduction in false detection rate achieved by the second.

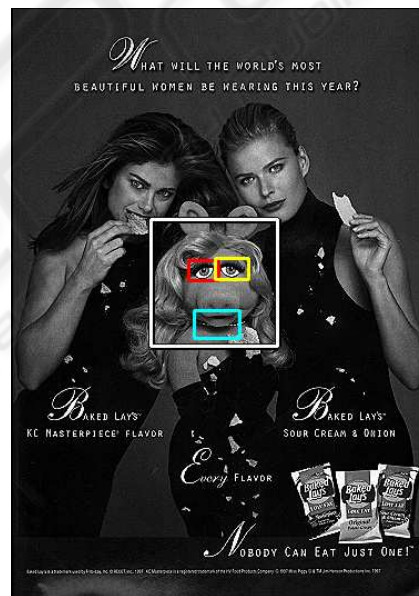


Figure 7: Example of false face detection that reported at least two inner facial features corresponding to *rot5*. image (contained in CMU dataset (Schneiderman and Kanade, 2000)) from the *rotated* subset. Does the reader consider it a false detection? Other detections are present with *soccer* and *Germany* images, but not included here due to space restrictions.

4 CONCLUSIONS

In this paper we have reviewed the possibilities of current public domain classifiers, based on the Viola-Jones' general object detection framework (Viola and Jones, 2004), for face and facial feature detection. We have observed a similar performance achieved by three of those frontal face detectors included in the

OpenCV release. We have also compared some facial feature detectors available thanks to the OpenCV community, observing that they perform worse than those designed for frontal face detection. However, this aspect can be justified by the lower resolution of the patterns being searched.

Finally we observed the possibilities provided by a simple combination of those classifiers, applying coherently facial feature detection only in those areas where a face has been detected. In this sense, the facial feature classifiers can be applied with more detail without remarkably increasing the processing cost. This approach performed better, improving facial feature detection and reducing the number of false positives. These results have been achieved without any further restriction in terms of facial feature cooccurrence, relative distances and so on. Therefore, we consider that further work can be done to get a more robust cascade approach using the public domain available classifiers, providing reliable information.

We consider also that the combination of face and facial feature detection can improve face detection, adding reliability to the traditional face detection approaches. However, the solution requires more computational power due to the fact that more processing is performed.

ACKNOWLEDGEMENTS

Work partially funded the Spanish Ministry of Education and Science and FEDER funds (TIN2004-07087).

REFERENCES

- Bediz, Y. and Akar, G. B. (2005). View point tracking for 3d display systems. In *3th European Signal Processing Conference, EUSIPCO-2005*.
- Bradley, D. (2003). Profile face detection. <http://www.davidbradley.info/publications/bradley-iurac-03.swf>. Last accessed 5/11/2007.
- Carnegie Mellon University (1999). CMU/VACS image database: Frontal face images. http://vasc.ri.cmu.edu/idb/html/face/frontal_images/index.html. Last accessed 5/11/2007.
- Castrillón Santana, M., Déniz Suárez, O., Hernández Tejera, M., and Guerra Artal, C. (2007). ENCARA2: Real-time detection of multiple faces at different resolutions in video streams. *Journal of Visual Communication and Image Representation*, pages 130–140.
- Intel (2006). Intel Open Source Computer Vision Library, v1.0. <http://sourceforge.net/projects/opencvlibrary/>.
- Jesorsky, O., Kirchberg, K. J., and Frischholz, R. W. (2001). Robust face detection using the hausdorff distance. *Lecture Notes in Computer Science. Procs. of the Third International Conference on Audio- and Video-Based Person Authentication*, 2091:90–95.
- Kruppa, H., Castrillón Santana, M., and Schiele, B. (2003). Fast and robust face finding via local context. In *Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, pages 157–164.
- Li, S. Z., Zhu, L., Zhang, Z., Blake, A., Zhang, H., and Shum, H. (2002). Statistical learning of multi-view face detection. In *European Conference Computer Vision*, pages 67–81.
- Liang, L., Liu, X., Pi, X., Zhao, Y., and Nefian, A. V. (2002). Speaker independent audio-visual continuous speech recognition. In *International Conference on Multimedia and Expo*, pages 25–28.
- Lienhart, R., Kuranov, A., and Pisarevsky, V. (2003a). Empirical analysis of detection cascades of boosted classifiers for rapid object detection. In *DAGM'03*, pages 297–304, Magdeburg, Germany.
- Lienhart, R., Liang, L., and Kuranov, A. (2003b). A detector tree of boosted classifiers for real-time object detection and tracking. In *IEEE ICME2003*, pages 277–80.
- Lienhart, R. and Maydt, J. (2002). An extended set of haar-like features for rapid object detection. In *IEEE ICIP 2002*, volume 1, pages 900–903.
- Pentland, A. (2000). Looking at people: Sensing for ubiquitous and wearable computing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, pages 107–119.
- Reimondo, A. (2007). Haar cascades repository. <http://alereimondo.no-ip.org/OpenCV/34>.
- Rowley, H. A., Baluja, S., and Kanade, T. (1998). Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):23–38.
- Schneiderman, H. and Kanade, T. (2000). A statistical method for 3d object detection applied to faces and cars. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1746–1759.
- Sung, K.-K. and Poggio, T. (1998). Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(1):39–51.
- Urtho (2006). Eye detector. <http://face.urtho.net/>. Last accessed 5/9/2007.
- Viola, P. and Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, 57(2):151–173.
- Wilson, P. I. and Fernandez, J. (2006). Facial feature detection using haar classifiers. *Journal of Computing Sciences in Colleges*, 21:127–133.
- Wimmer, M. (2004). Eyefinder. <http://www9.cs.tum.edu/people/wimmerm/se/project.eyefinder/>. Last accessed 5/11/2007.