

Eye Center Localization Using Adaptive Templates

J. Rurainsky and P. Eisert

Fraunhofer Institute for Telecommunications - Heinrich-Hertz Institute
{rurainsky,eisert}@hhi.fraunhofer.de

Abstract

We present a reliable eye center localization algorithm from one color image of a human face. The detection is based on deformable templates, which are only controlled by four parameters. Our algorithm works instantaneously and does not need a preprocessing step like the creation of a database or initial parameters. The experimental results show that the presented technique is very robust under different illumination conditions including pictures of people with glasses and different face sizes.

1. Introduction

Face features can be used in a variety of applications like face recognition, view point approximation, and in a communication system based on facial animation parameters, which are applied to a 3D head model. Such a communication system usually requires an initial detection of feature points, which are used to adjust a generic 3D head model in shape and orientation to the individual captured in the first frame. The requirements of detected face features for a communication system, like the one described above, are the accuracy of the detected positions as well as the amount of face features.

The detection of facial features has been approached by many researchers and a variety of methods exist. Nevertheless, because of the complexity of the problem, the robustness and necessary preprocessing steps of these approaches are still a problem. Most commonly, natural face feature templates taken from real persons are used for a template matching algorithm [1][2][3][4][5]. These templates have to satisfy a set of requirements like orientation, size, and illumination. Therefore a preprocessing step is necessary to align and size the templates. A wavelet-based approach, in which face images and face features from a database have to be aligned in orientation and size in a preprocessing step, is described in [6]. Both previous described methods are limited by the used templates and face database. Deformable templates also belong to the class of artificially constructed templates. This approach is using more detailed templates for eye and mouth outline matching and needs initial parameters [7]. Non realistic templates are used in this method

[8]. Another approach uses an iterative luminance threshold for the search window and geometric constraints [9].

We propose a deformable template approach for the localization of eye centers using one head and shoulder picture. Our approach does not need a preprocessing step like the creation of a natural face feature template database or initial parameters, which is the major difference considering that such a database more or less limits the range of usage to the templates included in the database. Another advantage of our approach are the simple adaptive templates, which can be controlled by only four position parameters. This small number of parameters limits the range of changes and therefore the number of possible template shapes.

2. System Description

Our algorithm can be split into two parts. First, we reduce the processing time through determination of the region of interest. In the case of eye center detection, this region of interest is a region around the eyes. Second, we describe the eye center detection as well as the used templates in detail.

Fig. 1 illustrates the two steps. Please, note that the system utilizes face features relations in all processing steps, like the relations described in [10].

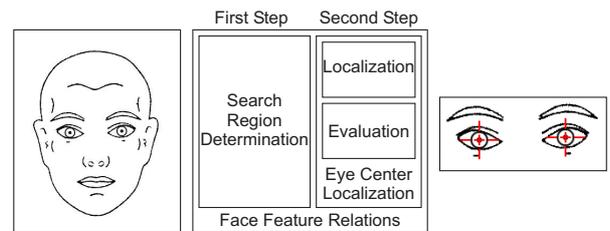


Figure 1: (left) Input picture of a human face. (middle) Eye center detection system. (right) Detected eye center pair.

3. Face Feature Relations

With **Fig. 2** the face feature measurements are given, which are used for eye center localization. The length of the eye fissure is corrected by the inclination of the eye fissure given by [10] as well.

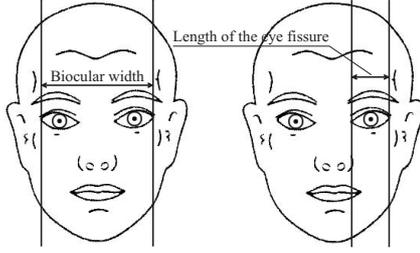


Figure 2: Face feature measurements given by [10]. The solid line represents the mean measurements of the specific face feature.

The relations of the face features can be extracted from the given anthropometric data for example the relation of the eye width (length of the eye fissure) to the eye distance (biocular width). This relation can be set to approximately:

$$\frac{\text{Biocular width}}{\text{Length of the eye fissure}} \approx \frac{1}{\frac{1}{3}}$$

4. Eye Center Localization

The algorithm has two major parts, which will be explained more in detail (cost functions, face feature relations) in the following sections. We start with a brief description of our algorithm before giving more details.

Our method starts to narrow the search region using a face detection algorithm [11]. This step locates the face within the image. Next a subset of templates from the template space is used to determine the appropriate horizontal gradient relations and the degree of color separation between the skin and iris-pupil region. This step is used as well as the previous step to minimize the possible search region. Given the search region, the eye center localization is performed using the full template space. The templates, which show the best match on a specific search position are used to separate the skin from iris-pupil region by computing a weight matrix. This weight matrix excludes low localization results. Analyzing the highest values of the weighted localization results leads to a set of possible eye positions. Face feature relations help to extract the desired eye pair.

In the following we separate our localization algorithm in two parts. First, the "Search Region Determination", which describes the face detection, horizontal gradient and color separation determination using a subset of the available templates. Second, the "Eye Center Localization", which performs a full search with all available templates, and the evaluation of the search result.

4.1. Search Region Determination

In a first step, the face is localized within the given image. We are using the IIS Real-time face detection library for this step [11]. The result of this algorithm are two positions close to the eyes of the person shown within the image. The deviation between the face detection results and the real eye positions are given in **Fig. 3**. Eye widths of the used face detection method are highly correlated to the real eye widths (ecc of 0.8). The mentioned empirical correlation coefficient (ecc) as shown in **Fig. 3** describes the degree of correlation and is defined between minus one and one, where zero means no correlation.

$$\text{ecc} = \frac{s_{xy}}{s_x \cdot s_y}$$

The empirical covariance s_{xy} and the square root of the bias-corrected variances (standard deviations) s_x and s_y are used to determine the ecc.

Because of the adaptive face templates used by the face detection library (five face templates are shown in **Fig. 3**), this method can be used for a large range of face sizes.

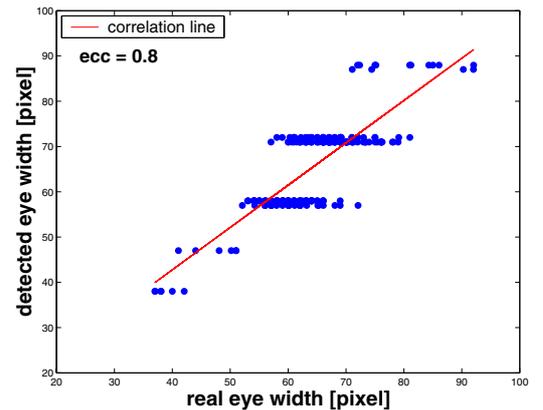


Figure 3: Correlation between the IIS Real-time detection library determined vertical eye widths and the vertical eye widths of the real eye positions.

The distances of the determined positions from the real eye positions (left and right eye separately) given with **Fig. 4** show almost no correlation. This allows to define a search region independent of the face size (number of skin colored pixels). Our evaluation of the IIS Real-time face detection library showed a 98.6% reliability. This reliability is determined by the useful specified eye positions. Useful face detection results return two positions close to the eye center positions. Negative labeled test results were excluded, in order to evaluate our eye center detection approach independently from the IIS Real-time face detection library results.

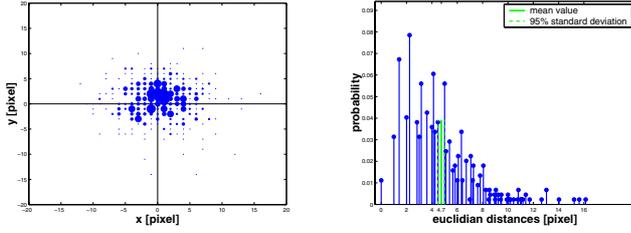


Figure 4: (left) Euclidean distances between the face detection results and the real eyes for the left and right eye measured in pixel. (right) Distribution histogram of the results shown left.

The search region can be fixed to a specific number of pixels around the face detection results, because of the high correlation between this result and the real eye positions. Furthermore, the face detection result is used in a linear equation to determine the initial template iris diameter. Around this initial iris diameter the template space is created. Face feature relations are given in section 3: "Face Feature Relations". The result of this step are search regions around the face detection results covering the real eyes, and the initial template iris diameter.

The dark iris-pupil disk region is surrounded by bright regions. These boundaries conditions of the iris-pupil disk to the eyeball (white colored) and to the lower lid (skin colored) can be used to establish useful eye properties. These properties are used in a second step of the search region detection method. The appropriate horizontal gradient relations and values are detected. Appropriate horizontal gradients have opposed orientation along the left and right iris to eyeball boundary, and show high absolute values. This evaluation requires templates, which are a subset of the template space. The used templates lost the freedom to perform a horizontal deformation. Only vertical deformations are allowed. Finding possible template iris diameters and to exclude search regions is the purpose of this step. The used cost function for the templates is specified as follows:

$$CFM(x, y, ri-p, le-p) = \sum_{(i,j) \in right}^{N_{right}} \frac{\delta I(i, j)}{\delta x} - \sum_{(k,l) \in left}^{N_{left}} \frac{\delta I(k, l)}{\delta x}.$$

The subscripts *left* and *right* are specifying the sub templates of the iris outline template shown in **Fig. 6**. The positions parameters *ri-p* and *le-p* are associated the used sub templates.

The possible colors of the iris-pupil disk are found in a different region of the RGB color space than the skin. Therefore a plane can be specified to separate these both

color regions. The third and last step of minimizing the search region uses the found template positions from the previous step to determine the skin to iris-pupil color separation degree, which is the distance from the mentioned plane. This color evaluation step provides us with a very small search region and boost therefore the following pixel-wise eye center localization. In order to determine the desired color separation degree the iris sub template shown in **Fig. 7** is used together with a cost function. This cost function is specified as follows:

$$DOSM(x, y) = \frac{1}{N_{iris}} \cdot \sum_{i=1}^{N_{iris}} [(\vec{p}_i(x, y) - \vec{p}_{plane}) \cdot \vec{n}_{plane}].$$

There \vec{p} is a point in the RGB space on the color separation plane, and \vec{n} is the normal on this plane showing towards the iris color region. The result of this step is the degree of separation matrix (DOSM) and is used as a weight matrix. This weight matrix is applied to the cost function matrix (COM) of the search region detection step.

$$SearchRegion = DOSM_{normalized}^{(+)} \cdot CFM_{normalized}^{(+)}$$

The result of the matrix multiplication defines the desired search region.

4.2. Localization

Given the search region described in section 4.1 the eye centers are determined, which are in the middle of the iris-pupil disk. This search consists of two steps. First, eye center localization within the predefined search region. Second, evaluation of the localization results.

The first step is described by a pixel-wise search using all different deformed templates of the template space together with a cost function at each search position. The cost function for the eye center detection has the following parts:

$$\begin{aligned}
& CFM(x, y, ri-p, le-p, lo-p, up-p) \\
= & \sum_{(i,j) \in right}^{N_{right}} \frac{\delta I(i, j)}{\delta x} - \sum_{(k,l) \in left}^{N_{left}} \frac{\delta I(k, l)}{\delta x} \\
+ & \sum_{(m,n) \in lower}^{N_{lower}} \frac{\delta I(m, n)}{\delta y} - \sum_{(o,p) \in upper}^{N_{upper}} \frac{\delta I(o, p)}{\delta y} \\
+ & CORR_{horizontal} \\
+ & CORR_{vertical} \\
- & \left| \sum_{(q,r) \in upper2}^{N_{upper2}} \frac{\delta I(q, r)}{\delta x} - \sum_{(s,t) \in lower2}^{N_{lower2}} \frac{\delta I(s, t)}{\delta x} \right| \\
= & \frac{CORR(x, y)}{\sqrt{\frac{1}{M} \sum_{i=1}^M (a_i - m_a)^2 \cdot \frac{1}{M} \sum_{i=1}^M (a_i - m_a)^2}}.
\end{aligned}$$

Subscripts *left*, *right*, *upper* and *lower* are specifying the sub templates of the iris outline template shown in **Fig. 6**. The subscripts *horizontal* and *vertical* are used to label the usage of the horizontal and vertical templates shown in **Fig. 7**. The subscripts *upper2* and *lower2* are the upper and lower sub templates of the horizontal and vertical templates.

The elements of the eye center localization cost function are horizontal gradients for the iris-eyeball boundary using the left and right sub templates shown in **Fig. 6**, vertical gradients for the iris-eyelid boundary using the upper and lower sub templates shown in **Fig. 6**, correlations of sub images with the horizontal and vertical templates as well as the horizontal gradients using the second version of the upper and lower sub templates shown in **Fig. 7**. In order to determine the horizontal and vertical correlation results the luminance values of the iris-pupil disk and the surrounding sub template areas are estimated for each search position in order to minimize the difference of synthesized horizontal as well as vertical templates to the search image. Each element of the cost function has an equal weight in the cost function.

4.3. Evaluation

This step evaluates and groups localization results, in order to determine eye center position pairs. First, the degree of color separation between the iris-pupil disk and a set of surrounding sub templates is determined. The desired color information are extracted using the previously determined templates. The color separation plane together with a default template for this step are shown in **Fig. 5**.

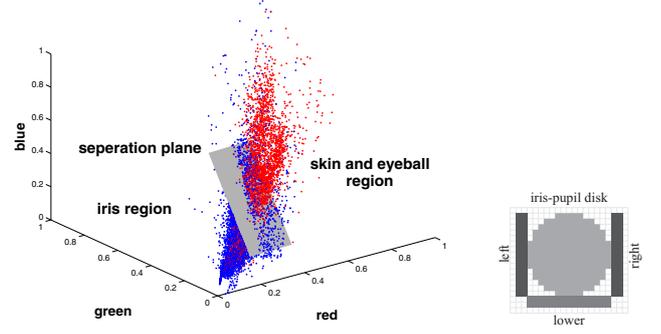


Figure 5: (left) Iris-skin/eyeball color separation plane. (right) Used templates for determination of the separation degree.

The idea behind this approach is, that the left and right sub templates extract eyeball color information and the lower sub template extracts skin color information, if the sub templates show the right shape and position. Accumulating the distances to the plane leads to the degree of separation. The used cost function is described as follows:

$$\begin{aligned}
& DOSM(x, y) \\
= & \frac{1}{N_{iris}} \cdot \sum_{i=1}^{N_{iris}} [(\vec{p}(x, y) - \vec{p}_{plane}) \cdot \vec{n}_{plane}].
\end{aligned}$$

There \vec{p} is a point in the RGB space on the color separation plane, and \vec{n} is the normal on this plane showing towards the iris color region. The result of this step is the degree of separation matrix (DOSM) and is used as a weight matrix. This weight matrix is applied to the cost function matrix (COM) of the eye center detection step.

$$\begin{aligned}
& EyeCenterRegion \\
= & DOSM_{normalized}^{(+)} \cdot CFM_{normalized}^{(+)}
\end{aligned}$$

The result of the matrix multiplication defines the desired search region.

Applying face feature relations to the weighted detection matrix allows to define possible eye pairs. For this step statistical descriptions of face feature relations such as [10] and [12] are used. In section 3 the used face feature relations are described.

5. Templates

The templates which are used for the eye center localization algorithm are described in this section as well as the creation pyramid for deforming the default shaped templates.

Each template consists of several sub templates. Size and position are the parameters, on which the sub templates depend on each other.

The iris outline template, which is used in the eye center localization procedure consists of four sub templates and approximates the circular iris-pupil disk as a rectangle. The iris-pupil disk can be partly covered by the upper and lower eye lid. In order to take these occlusions into account, the upper and lower eye lid are approximated as straight lines. These four sub templates are shown in **Fig. 6** and called left, right, upper and lower sub template. Each sub template has the capability to change the position independent of the other sub templates. Therefore four position parameters are used to control the iris outline template. These four parameters are **le-p** (left position), **ri-p** (right position), **up-p** (upper position) and **lo-p** (lower position). The size of the left and right sub template depends on all four position templates. The size of the upper and lower sub templates depends only on the position parameters of the left and right sub template.

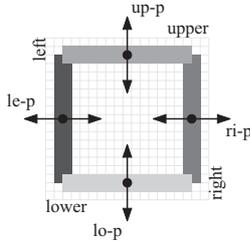


Figure 6: Iris outline template composed of left, right, upper and lower sub templates. Position parameters $le-p$, $ri-p$, $up-p$ and $lo-p$ are used to control this template.

The horizontal and vertical templates, which are used for correlation consist of five sub templates. These five sub templates are shown in **Fig. 7** and called iris-pupil disk, left, right, upper and lower sub template. In order to control these five sub templates four positions parameters are used. The same deformation rules as described for the iris outline template parameters are used for the horizontal and vertical template parameters.

Because of the dependency between the three described templates only four parameters are necessary to control all three templates and associated sub templates.

The previously described templates are default templates. In order to use these templates a template space has to be created, which holds different template shapes. The given template creation pyramid in **Fig. 8** shows all possible changes of the described templates.

The initial template is created by the diameter of the iris-pupil disk, which is determined by the distance of the detected eye positions of the IIS face detection algorithm and

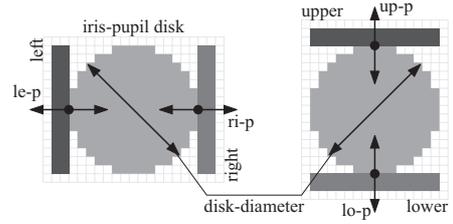


Figure 7: (left) Horizontal template composed of left, right and iris-pupil disk sub templates. (right) Vertical template composed of upper, lower and iris-pupil disk sub templates. Position parameters $le-p$, $ri-p$, $up-p$ and $lo-p$ are used to control this template.

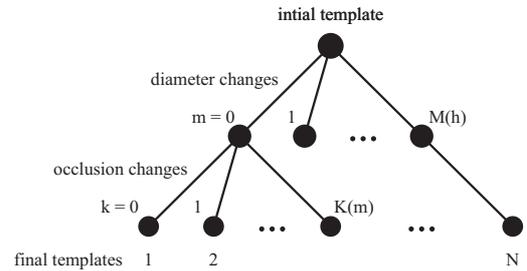


Figure 8: Template creation pyramid.

face feature relations such as [10]. Diameter changes are applied by the left and right position parameters. Occlusion changes are applied by the upper and lower position parameters.

Specific rules can be applied during template space creation in order to minimize the amount of different templates and to exclude non realistic templates. First, to avoid multiple templates at one detection position the maximum shift between the left and right sub templates is limited to one. Second, the iris-pupil disk occlusion of the upper and lower sub templates is limited to the half diameter of the current iris-pupil disk.

6. Experimental Results

We have applied the described eye center localization algorithm using adaptive templates to a data set of 410 head and shoulder images of 348 different people. The images originate from two different databases. One database holds 187 images with CIF (352x288 pixel) resolution and was captured by the HHI using a digital camera. The other database is a subset of the *FERET* database. Applying the same rules, like head orientation and eye opening, as shown in the HHI database allow to use 223 head and shoulder images of the *FERET* database. The taken images from the *FERET*

database have a resolution of 256x384 pixel, which is half of horizontal and vertical resolution provided by the *FERET* database. We reduced the provided resolution in this way, because of our target resolution, which is CIF. Reducing the pixel resolution still provides a large range of face sizes, which can be shown in **Fig. 2**. The combined database holds images with different illumination conditions (e.g. inside, outside, flash, neon light), different face/skin color, sizes, people with glasses, male, female and kids. There is one constraint common to all images: the person has to look straight into the camera. Examples of the HHI face database can be found in **Fig. 9**.



Figure 9: Examples of the HHI database.

The Euclidean Distance was used for the evaluation between manually selected eye centers and the found eye centers. The manual selected eye centers were determined with full pixel accuracy. The detection algorithm returns the eye center position with half pixel accuracy.

eye pair choice	detection evaluation [percentage]	
	HHI database	FERET database
first	95.2	95.5
second	2.7	0.4
third	0.0	0.0
total	97.9	95.9

Table 1: Reliability of the first three detected eye center pairs.

Tab. 1 shows the reliability of this detection algorithm. We have collected the first three possible eye pairs, because of the allowed range of face feature relations. In a following sequence analysis the second or third pair could be used as well. The algorithm shows a reliability of around 95% for both face databases. We have counted as positive detected eye pairs detection results within a three pixel reliable region around each manual selected eye center. Therefore the

maximum allowed Euclidean Distance of each detected eye center were limited to three pixel.

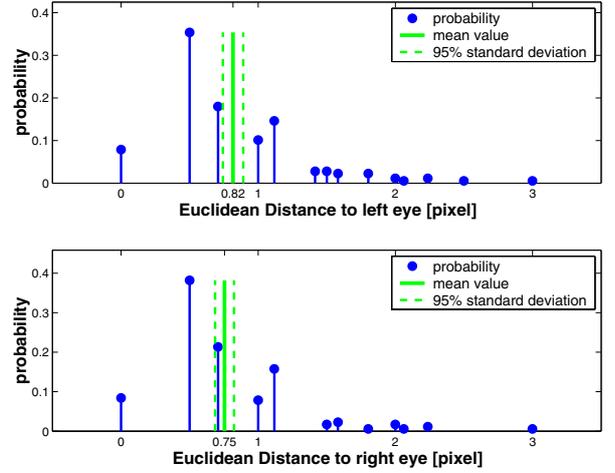


Figure 10: Evaluation of detected eye center positions by comparison with manual selected eye centers of the HHI face database. (top) Results for the left eye centers. (bottom) Results for the right eye centers.

The histograms **Fig. 10** and **Fig. 11** for the HHI and *FERET* database show the distribution of the Euclidean Distances as well as the mean value and 95% standard deviation for the left and right eye center separately. These histograms were created using only the first detected eye pair choice.

The average Euclidean Distance of both eye centers and both face databases is less than one pixel. Examples of the positive localized eye center are given in **Fig. 12**.

Examples of miss detections, which occur with 5% are shown in **Fig. 13**. Possible miss detections could occur in cases of thick glasses or eyebrows. In such a case another eye pair center choice could be used.

7. Summary and Conclusions

We have described a novel eye center localization algorithm with adaptive templates. The detection procedure is mainly based on gradient, luminance, and color information. The templates are controlled only by four position parameters.

The achieved result for the localization of eye centers shows a reliability of around 95% by an average accuracy of less than one pixel (Euclidean Distance). The major advantage of this algorithm is the very simple setup procedure. No database creation, like for an Eigen or Wavelet approach is needed. Another advantage is the lack of parameters range limitations, which have to be set either manually or determined through a trainig procedure.

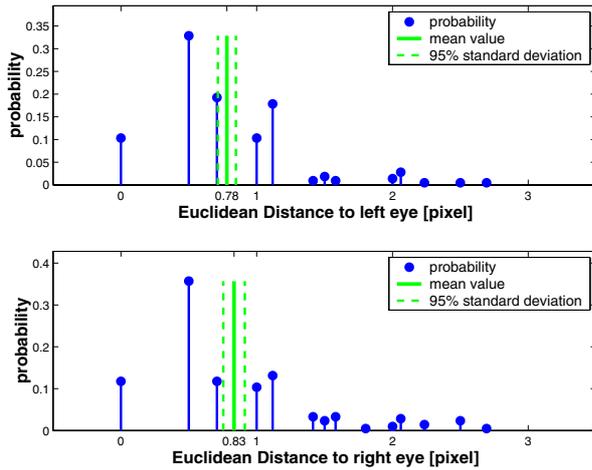


Figure 11: Evaluation of detected eye center positions by comparison with manual selected eye centers of the *FERET* database. (top) Results for the left eye centers. (bottom) Results for the right eye centers.



Figure 13: Eye center detection results of the HHI database. Results outside the range of three pixel Euclidean Distance from manual selected positions.

The current implementation is not optimized for real-time processing, which is subject to future research. Other future steps will include a change from the template pyramid to a hierarchical approach for the template creation.

Acknowledgments

The work presented in this paper has been developed with the support of the European Network of Excellence VISNET (IST Contract 506946).

A special thank to Christian Küblbeck of the Fraunhofer Institut Integrierte Schaltungen for providing the Real-time face detection library, which is used in the initial face segmentation steps.

In order to test our algorithm we used portions of the *FERET* database of facial images collected under the *FERET* program, therefore a special thank to the National Institute of Standards and Technology (NIST) for providing this database [13].

References

- [1] Markus Kampmann, "Automatic 3-D Face Model Adaptation for Model-Based Coding of Videophone Sequences," *IEEE Transaction on Circuits and Systems for Video Technology*, vol. 12, no. 3, Mar. 2002.
- [2] Rogério Schmidt Feris, Teófilo Emídio de Campos, and Roberto Marcondes Cesar Junior, "Detection and Tracking of Facial Features in Video Sequences," in *Proceedings of the Mexican International Conference on Artificial Intelligence (MICA I 2000)*, Acapulco, Mexico, April 2000, vol. 1793 of *Lecture Notes in Computer Science*, Springer.
- [3] Markus Kampmann and Jörn Ostermann, "Automatic adaptation of a face model in a layered coder with an object-based analysis-synthesis layer and a knowledge-based layer," *Signal Processing: Image Communication*, vol. 9, pp. 201–220, 1997.
- [4] Lars-Peter Bala, Kay Talmi, and Jin Liu, "Automatic Detection and Tracking of Faces and Facial Features in Video Sequences," in *Proceedings of the International Picture Coding Symposium (PCS 1997)*, Berlin, Germany, September 1997, pp. 251–256.
- [5] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI 1993)*, vol. 15, no. 10, pp. 1042–1052, October 1993.
- [6] R. S. Feris, J. Gemmell, K. Toyama, and V. Krueger, "Hierarchical Wavelet Networks for Facial Feature Localization," in *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, Washington D.C., USA, May 2002.
- [7] A. L. Yuille, "Deformable Templates for Face Recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 59–79, 1991.
- [8] Jürgen Rurainsky and Peter Eisert, "Template-based Eye and Mouth Detection for 3D Video Conferencing," in *Proceedings of the International Workshop on Very Low Bitrate Video (VLBV 2003)*, Madrid, Spain, September 2003, pp. 23–31.
- [9] Jie Yang, Rainer Stiefelhagen, Uwe Meier, and Alex Waibel, "Real-Time Face And Facial Feature Tracking And Applications," in *Proceedings of the Workshop on Audio-Visual Speech Processing (AVSP 1998)*, Terrigal, South Wales, Australia, December 1998, pp. 79–84.
- [10] Leslie G. Farkas, *Anthropometry of the Head and Face*, Raven Press, 2nd edition, 1995.
- [11] Bernhard Fröba, Andreas Ernst, and Christian Küblbeck, "Real-time face detection," in *4th IASTED International Conference on Signal and Image Processing (SIP 2002)*, Kauai, August 2002, pp. 479–502.
- [12] Leslie G. Farkas and Ian R. Munro, *Anthropometric Facial Proportions in Medicine*, Charles C Thomas Pub Ltd, 1987.
- [13] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image and Vision Computing Journal*, vol. 16, no. 5, pp. 295–306, 1998.

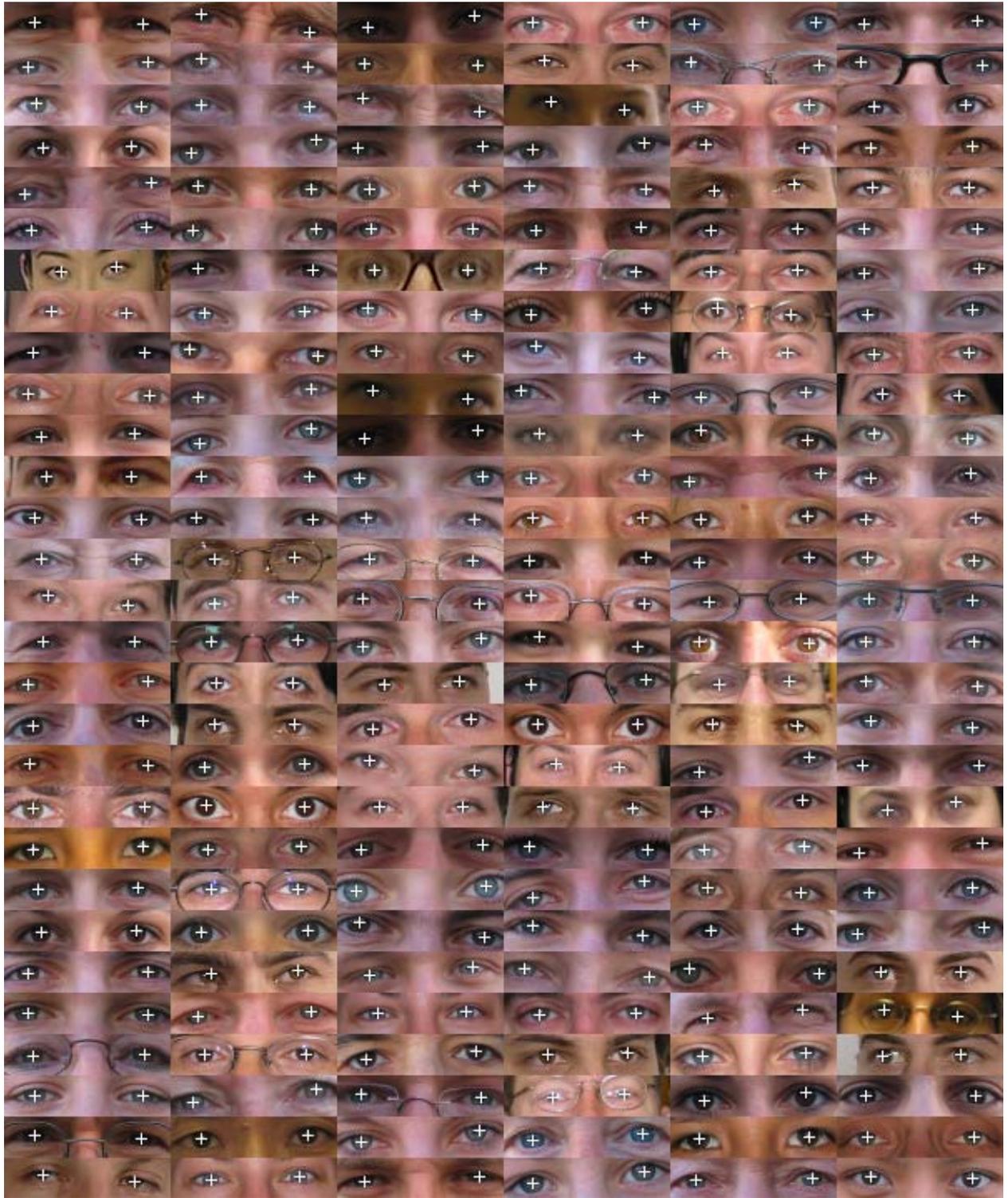


Figure 12: Eye center detection results of the HHI database. Results within three pixel Euclidean Distance from manual selected positions.