THREE-DIMENSIONAL HUMAN FACE FEATURE EXTRACTION FROM MULTI IMAGES

Nicola D'APUZZO* **and Nobuo KOCHI**[§]
* Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland
nicola@geod.baug.ethz.ch
§ Imaging Laboratory, R&D Center, TOPCON Corporation, Tokyo, Japan
n.kochi@topcon.co.jp

KEY WORDS: Human Face, Feature Extraction, Progressive Scan, IEEE-1394.

ABSTRACT

In this paper, we describe methods developed and preliminary results achieved by a fully automatic system for the measurement of 3-D face features using multi-images acquired simultaneously by multiple cameras. The method is based on multi-image photogrammetry and consists basically of five steps: calibration of the multiple camera system, acquisition of multi-images, edge extraction in all the acquired images, edge matching in the different images and computation of 3-D edges.

A multi-image acquisition system was especially designed for face measurement purposes. Five IEEE-1394 progressive scan CCD cameras arranged convergently are used. A graphical user interface was developed for the real-time camera control, for the real-time display of multiple camera views and for the synchronized acquisition of multi-images. A reference field with signalized points was especially designed for the five camera acquisition system to allow a full automatic calibration procedure. The process is integrated in the graphical user interface software and the unique task required for the user is to place the calibration field at the indicated position. The signalized points are then fully automatically recognized and measured in the different images and the establishment of the external and internal orientation as well as parameters modeling the lens distortion is performed by bundle adjustment.

As preprocessing of the face feature extraction, smoothing and edge extraction is performed on the different images independently. A combination of SUSAN smooth and Canny edge extraction resulted to be the most effective and robust solution for the application on human faces. To increase the speed of the entire process, multiple regions of interest can be defined manually in the images.

The full automatic edge matching process is based on the geometrical constrained least squares matching algorithm. Its task consists of the establishment of correlations between the edges extracted in the different images. The edge matching process uses one image as template and the others as search images. Each edge point extracted in the template image is defined as template point and its corresponding points are searched in the search images along the epipolar lines. As candidates, only extracted edge points are used. To evaluate the quality of the matching results, standard deviation of the least squares adjustment process is used and to detect possible mismatches, the distance to multiple epipolar lines is checked. The last step of the process is the computation of the 3-D coordinates of all the matched edge points by forward ray intersection, resulting in a set of 3-D edges of human face features.

This papers present the preliminary results achieved in the newly started project. The result achieved by the implemented method are robust enough. However, the time required by the matching process is not yet satisfying our final requirements.

1 INTRODUCTION

Recently, many private and public research groups invested lot of time and resources in the development of face recognition systems for automatic person identification products. The dramatic events

occurred in New York two years ago accentuated even more the increasing demand of robust system for the identification of persons from image data. The majority of the methods used in commercial system as well as methods still under development that can be found in the literature, are based on single images. We suggest rather to utilize 3-D data extracted from multiple images because we are convinced that the robustness of face recognition systems would increase.

Facial feature extraction techniques are used by two major groups: the first deals with automatic face detection and tracking (Reinder et al., 1995, Sobottka and Pitas, 1997, Tian and Bolle, 2001, Feris et al., 2002, Froba and Kublbeck, 2002) and the second with face recognition (Turk and Pentland, 1991, Cox et al., 1996, Lincoln and Clark, 2001, Sakamoto et al., 2002).

In face detection and tracking different methods are used to date. Reinder et al. (1995) detect first the location and orientation of the face in the image and then extract some important features of the face (chin silhouette, eyes, nostrils, mouth) that are used for fitting a predefined 3-D face model. Feris et al. (2002) present a method based on hierarchical wavelets. A database of faces with stored features location is searched to find the best fit with the imaged face. The search is performed within wavelets. The location of the features is then refined using wavelets of each feature. Sobottka and Pitas (1997) use segmentation techniques to first localize the face in color images, then fit an ellipse to refine the results. The extraction of face features (mouth, eyes, eyebrows) is performed by using only the difference of intensity. The system is robust but not precise. Tian and Bolle (2001) present a method to detect different expression of faces. For this purpose, the eyes, eyebrows and mouth shapes have to be extracted from the acquired images. The face is first detected with image segmentation, it is then normalized to a predefined size. To extract the eyes position and shape, the image is binarized with a threshold and an ellipse is fitted to the eyes. Knowledge of the location of the eyes is used for the starting position. For the extraction of the mouth shape, a classical edge extractor is applied on the image, the mouth region is divided in 3x3 zones and the mouth shape is characterized as histogram of the edge map in the different zones. Froba and Kublbeck (2002) present a method to localize human faces based on the edge orientation field. A normalized set of human faces is used to define a dataset. The simple Sobel edge extractor is used to compute the edge map.

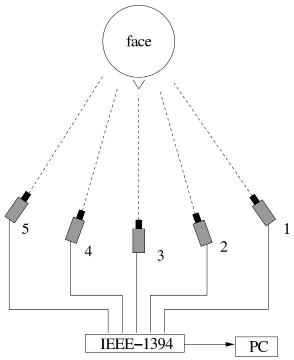
In face recognition a wide variety of methods can be found in the literature. Two different approaches can be distinguished: feature based approaches (Cox et al., 1996) which rely on a feature set and direct image methods (Turk and Pentland, 1991, Lincoln and Clark, 2001, Sakamoto et al., 2002) which involves no intermediate feature extraction stage. Cox et al. (1996) present a feature based method that involves 30 distances derived from 35 measured anatomical locations in frontal or nearly frontal views. The paper is focused on the recognition algorithm, therefore the 35 locations where measured manually. The recognition rate of the presented method reach 95% on a database of 685 people. NEC Multimedia research laboratories (Sakamoto et al., 2002) presented a 3-D model based face recognition method. They use a 3-D database to recognize faces in images. The images are compared with each registered 3-D shape and surface reflectance. Turk and Pentland (1991) uses eigenfaces in the following way: the face images are transformed in a small set of characteristic feature images (the eigenfaces) which are the principal components of the initial training set of face images. Recognition is performed by backprojecting a new image into the subspace spanned by the eigenfaces (face space) and then classifying the face by comparing its position in face space with the position of known individuals. Lincoln and Clark (2001) present a recognition system based on principle component analysis. The method involves taking each image in a video sequence, tracking the face from frame to frame and determining the head orientation in each frame, then merging the appropriate region of the image into a unwrapped texture map. Single frame of a person's head can be compared with a portion of the unwrapped texture map stored in the database.

In this paper, we describe methods developed and preliminary results achieved by a fully automatic system for the measurement of 3-D face features using multi-images acquired simultaneously by multiple cameras. The method is based on multi-image photogrammetry and consists basically of five steps: calibration of the multiple camera system, acquisition of multi-images, edge extraction in all the acquired images, edge matching in the different images and computation of 3-D edges.

2 ACQUISITION SYSTEM AND PREPROCESSING

2.1 Image acquisition

A multi-image acquisition system based on five IEEE-1394 progressive scan CCD cameras was especially designed for face measurement purposes. The five cameras mounted on a tripod are arranged convergently. A C-mount lens with 16 mm focal lens was chosen in order to compact the size of the complete system to about 1 m wide. The lens are provided with variable iris and the focus can be fixed. Figure 1 shows the disposition of the cameras and the setup arrangement. The entire system is mounted on a tripod that can move vertically and rotate around its vertical axis. The five cameras can rotate around two axes. Additionally, for the fine tuning of the system, the two extern cameras (number 1 and 5) can be displaced sideways and to the front or back. The central camera (number 3) can be moved to the front and back too. Three cameras (number 1, 3 and 5) are mounted on sockets to have a difference in the heights; in this way the epipolar lines will cross each others (see section 3.1).



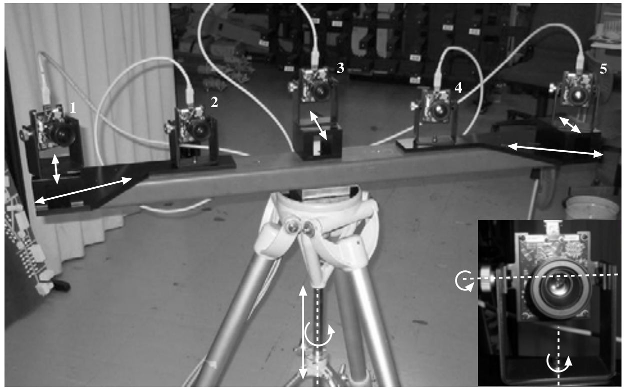


Figure 1: Multi-image acquisition system: disposition of the cameras and setup arrangement.

Figure 2 shows details of the chosen camera. It is a progressive scan CCD cameras (model *Dragonfly* of Point Grey Research) provided with a 1/3" sensor of 640x480 pixels and IEEE-1394 port. Through the IEEE-1394 interface, the computer communicates with the camera, allowing digital transmission of images as well as software control of some camera parameters. No frame grabber is therefore required. Different frame rates can be set from 3.75 Hz up to 30 Hz. Other features, such as brightness, exposure, gain, shutter and white balance can be controlled through the IEEE-1394 port. The Dragonfly comes equipped with the ability to perform automatically multiple camera synchronization. When more than one Dragonfly camera is present on the IEEE-1394 bus, the cameras automatically synchronize their acquisition time to within 20μs. This feature is particularly important for multi-image acquisition systems. A graphical user interface was developed for the real-time camera control, for the real-time display of multiple camera views and for the synchronized acquisition of multi-images.

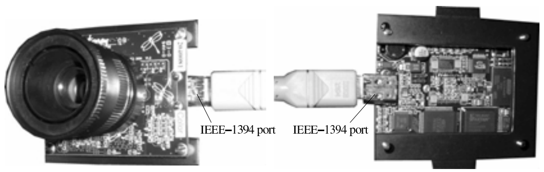


Figure 2: IEEE-1394 Progressive scan CCD camera *Dragonfly* from Point Grey Research.

2.2 Calibration and orientation

A reference field with signalized points was especially designed for the five camera acquisition system to allow a full automatic calibration procedure. The 3-D coordinates of the signalized points were measured photogrammetrically by *PI-3000* system (Kochi et al., 2003) of TOPCON with a precision of 0.2 mm.
The calibration and orientation processes are integrated in the graphical user interface software and the unique task required for the user is to place the calibration field at the indicated position. The signalized points are then fully automatically recognized and measured in the different images by least squares matching. The establishment of the external and internal orientation as well as parameters modeling the lens distortion is performed by bundle adjustment. A thorough determination of all these parameters is a requirement for a precise measurement from multi-images. Figure 3 shows the calibration field and the full automatic image coordinate measurement process as can be viewed in the graphical user interface.

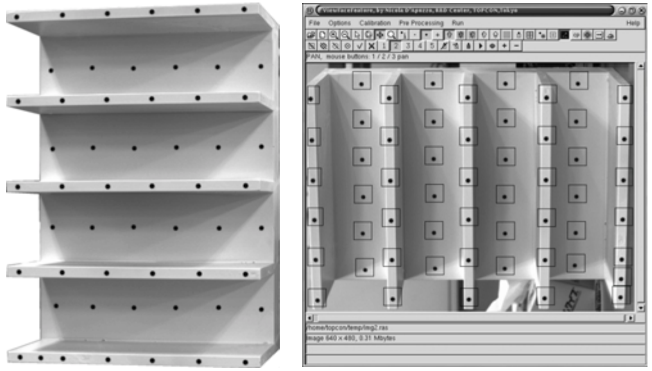


Figure 3: Left: calibration field. Right: full automatic measurement of signalized point.

2.3 Edge extraction

Several evaluations of edge extraction procedures were performed in the past years (e.g., Bowyer et al., 1999, Shin et al., 1999, Forbes and Draper, 2000), however different are the conclusions depending on the task to be performed and on the kind of object imaged. Forbes et al. (2000) also discuss the difficulties of this type of evaluation process. Therefore, it was decided to use standard edge extractors, e.g., Canny, SUSAN.

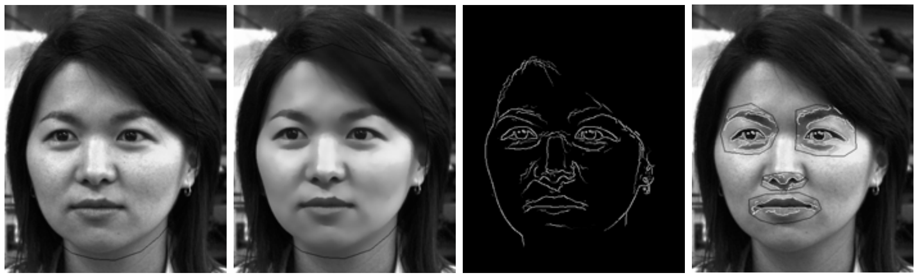


Figure 4: Preprocessing, from left to right: original image, SUSAN smooth, Canny edge operator, extracted edges in multiple ROIs.

As preprocessing of the face feature extraction, smoothing and edge extraction is performed on the different images independently. SUSAN and Canny edge extraction were implemented and tested with different parameters. A combination of SUSAN smooth and Canny edge operator resulted to be the most effective and robust solution for the application on human faces. To increase the speed of the entire process, multiple regions of interest can be defined manually in the images. This feature is especially important for the subsequent edge matching process which can be relatively time consuming. Figure 4 shows an original image and the different steps of image processing.

3 EDGE MATCHING AND 3-D COMPUTATION

3.1 Edge matching process

The task of the edge matching process is the establishment of correlations between the edges extracted in the different images. A full automatic method based on the geometrical constrained least squares matching algorithm (Gruen and Baltsavias, 1988, D’Apuzzo, 2002) was developed to solve this task. The edge matching process uses one image as template and the others as search images. Each edge point extracted in the template image is defined as template point and its corresponding points are searched in the search images along the epipolar lines in a restricted area (boxes in figure 5). As candidates, only extracted edge points are used. Least squares matching algorithm is applied at each position and the best matching results in considered the corresponding edge point. Figure 5 describes the process: three images (acquired by cameras 1, 2 and 3) are used, the image acquired by the camera number 2 is used as template image and the other two images as search images.

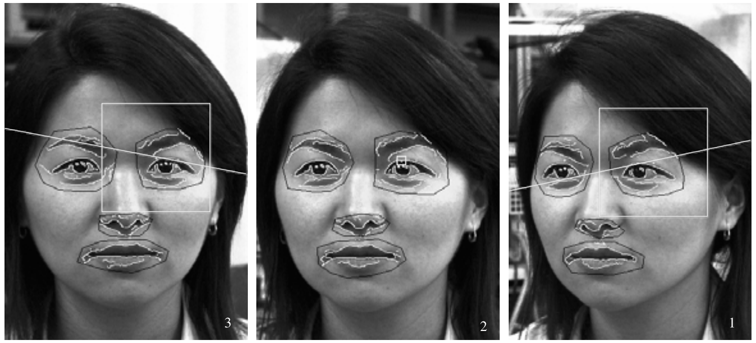


Figure 5: Edge matching. Center: template image, left and right: search images.

To evaluate the quality of the matching results, standard deviation of the least squares adjustment process is used and to detect possible mismatches, the distance to multiple epipolar lines is checked. Figure 6 shows an example of a matched edge point. The white boxes represent the image patches used by LSM, in the search images they are affinely transformed. The two lines represents the epipolar lines used for checking the result.

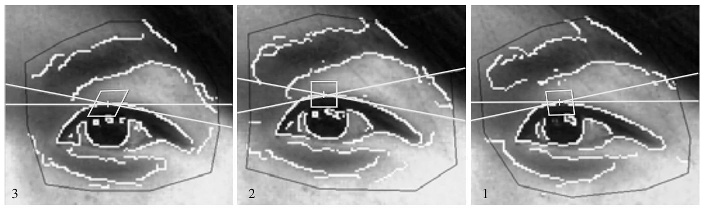


Figure 6: Edge matching result. Center: template image, left and right: search images.

3.2 3-D edges

The last step of the process is the computation of the 3-D coordinates of all the matched edge points by forward ray intersection, resulting in a set of 3-D edges of human face features. As the human face is a steep surface, the five images set is used as two triplets. The right side triplet is composed of the images acquired by the cameras number 5, 4 and 3 and the left side triplet by the images acquired by the cameras number 3, 2 and 1. Both triplets are treated independently resulting in two set of 3-D edges which are then merged together. No alignment process is required since the cameras are oriented absolutely. Figure 7 shows the edge matching results for the two triplets and figure 8 shows different views of the merged 3-D edges.

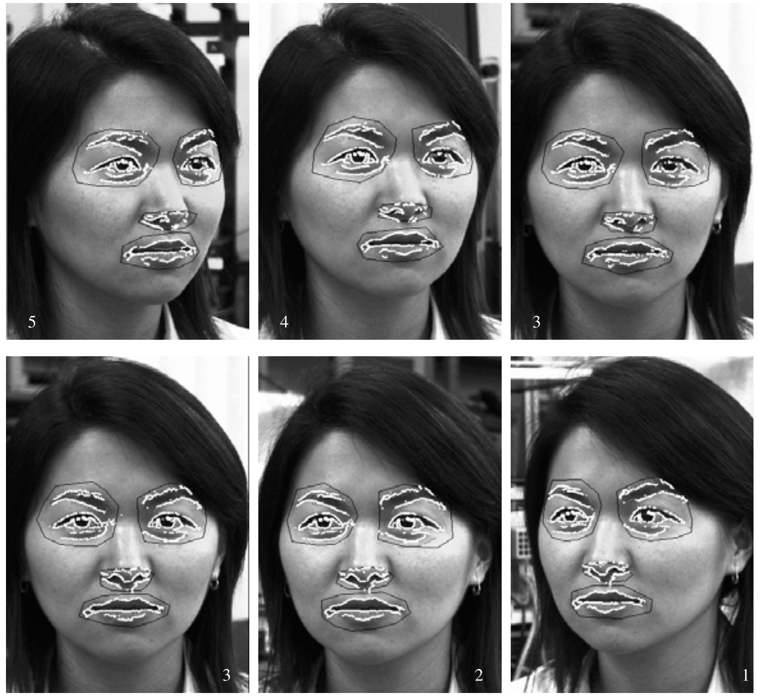


Figure 7: Edge matching result and 3-D edges. Top: right triplet, bottom: left triplet.

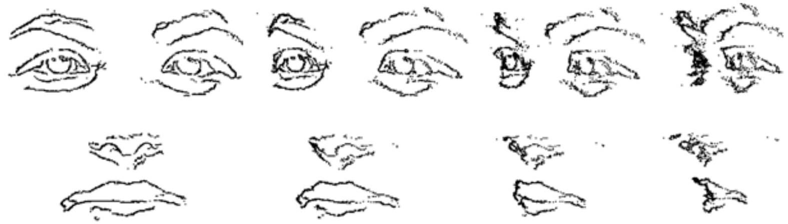


Figure 8: Different views of the merged 3-D edges.

The figure 9 shows in detail the extracted 3-D edges of the mouth. The different 3-D edges resulted from the left and right triplets can be recognized as distinguished polygon lines. The accordance of the two 3-D edges set is a sign of the good quality of the measurement achieved by the proposed method.

Figure 11 shows the extracted 3-D edges overlaid to the measured 3-D point cloud achieved by the method proposed by D’Apuzzo (2002) applied on the same images used for the edge extraction.

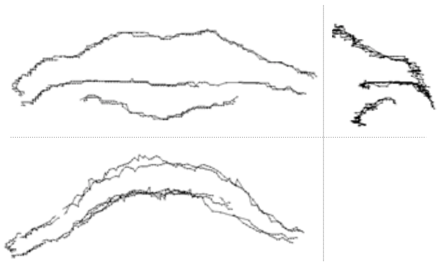


Figure 9: Detail mouth. Orthogonal views: front, side, bottom. The different 3-D edges resulted from the left and right triplets can be recognized as distinguished polygon lines.

4 CONCLUSION

This papers presents the preliminary results achieved in the newly started project. The result achieved by the implemented method are robust enough. However the time required by the matching process is not yet satisfying our final requirements. On a Pentium IV 2.0 GHz machine, the full automatic edge matching process requires from five to ten minutes, depending on the density of the extracted edges. This is a long processing time for person identification systems, which require the entire processing time to be kept under the minute. Strategies for reducing the processing time have therefore to be investigated for future developments on this project.

A second very important task to be further researched and implemented, is the automatic face feature identification from the set of measured 3-D edges. This would be the required input for a person identification process based on 3-D human face features that we believe would be more robust and have higher successful rate than systems based on single images. This assumption can be validated by the examples of figure 10, where the 3-D edges measured for ten persons (five females and five males) are shown. As can be observed in the orthogonal views, substantial differences are present in the position and shape of the three most important features of the human face: eyes, nose, mouth. Three-dimensional informations like 3-D angles, 3-D distances, 3-D curves could be measured in the set of 3-D edges and used for person identification.

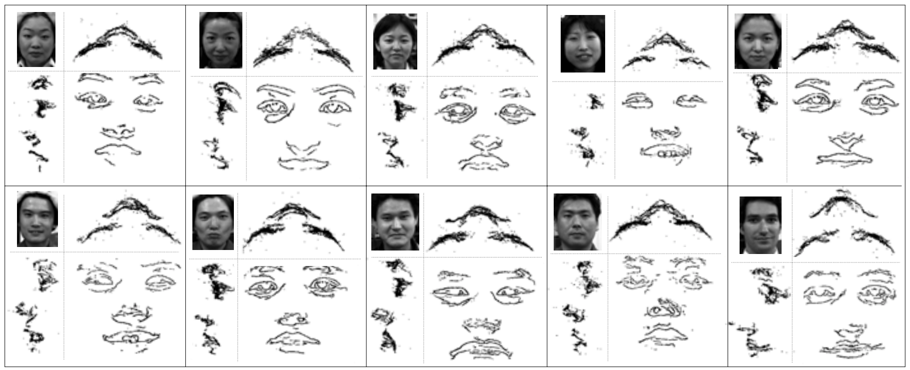


Figure 10: Examples. Central image and orthogonal views (top, side, front) of the extracted edges.

ACKNOWLEDGMENTS

The achievements shown in this paper are the result of seven months guest work of Nicola D'Apuzzo at the Imaging Laboratory of the R&D Center of TOPCON Corporation in Tokyo. We thank Professor emeritus Shunji Murai, Professor Yoshifumi Yasuoka and his laboratory of the Institute of Industrial Science of the University of Tokyo for the advise regarding this research and study. Thanks also to the friends and colleagues that posed for the acquisition of face data.

Figure 11: 3-D surface measurement and extracted 3-D edges.

REFERENCES

Bowyer, K., Kranenburg, C., Doughery, S., 1999: *Edge detector evaluation using empirical ROC curves.* In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 354-359.

Cox, I.J., Ghosn, J., Yianilos, P.N., 1996: *Feature-based face recognition using mixture-distance.* Proc. of International Conference on Computer Vision and Pattern Recognition, pp. 209-216.

D'Apuzzo, N., 2002: *Surface measurement and tracking of human body parts from multi image video sequences.* ISPRS Journal of Photogrammetry and Remote Sensing 56(5-6):360-375.

Feris, R.S., Gemmell, J., Toyama, K., 2002: *Facial feature detection using a hierarchical wavelet face database.* Technical Report MS-TR-2002-05, Microsoft Research, Redmond, USA.

Forbes, L.A. and Draper, B.A., 2000: *Inconsistencies in edge detector evaluation.* In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, Vol. 2, pp. 398-404.

Froba, B. and Kublbeck, C., 2002: *Robust face detection at video frame rate based on edge orientation features.* Face and Gesture Recognition 5:327-332.

Gruen, A. and Baltsavias, E.P., 1988: *Geometrically constrained multiphoto matching.* Photogrammetric Engineering and Remote Sensing, 54(5):633–641.

Kochi, N., Ito, T., Noma, T., Otani, H., Nishimura, S., Ito, J., 2003: *PC-based 3D image measuring station with digital camera. An example of its actual application on a historical ruin.* In Proc. of ISPRS WG V/4 & IC III/V Workshop, Ancona, Italy.

Lincoln, M.C. and Clark, A.F., 2001: *Pose-independent face identification from video sequences.* In Bigun, J., Smeraldi, F. (Eds.): Audio- and Video-Based Biometric Person Autentification, Springer, pp. 14-19.

Reinders, M.J.T., van Beek, P.J.L., Sankur, B., van der Lubbe, J.C.A., 1995: *Facial feature localization and adaptation of a generic face model for model-based coding.* Signal Processing: Image communication 7:55-74.

Sakamoto, S., Ishiyama, R., Tajima, J., 2002: *3D model-based face recognition system with robustness against illumination changes.* NEC Research & Development 43(1).

Shin, M.C., Goldgof, D., Bowyer, K.W., 1999: *Comparison of edge detectors using an object recognition task.* In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 360-365.

Sobottka, K. and Pitas, I., 1997: *Looking for faces and facial features in color images.* Pattern Recognition and Image Analysis 7(1).

Tian, Y.-L. and Bolle, R.M., 2001: *Automatic neutral face detection using location and shape features.* IBM Research Report RC 22259 (W0111-073).

Turk, M. and Pentland, A., 1991: *Eigenfaces for recognition.* Journal of Cognitive Neuroscience 3(1):71-86.