

# Low-cost face biometry for visually impaired users

L. Balduzzi, G. Fusco, F. Odone  
DISI, Università di Genova  
odone@disi.unige.it

S. Dini  
Istituto Chiossone  
silvia.dini@chiossone.it

M. Mesiti  
DiCo, Università di Milano  
mesiti@dico.unimi.it

A. Destrero, A. Lovato  
Imavis srl  
{destrero,lovato}@imavis.com

**Abstract**—We present a work in progress on a face biometry system for visually impaired users – the result of a very close interaction among scientists, engineers, and a users group formed by visually impaired and social assistants. The prototype under development implements recent trends of video analysis and follows closely the suggestions given by the working group, with the ambitious goal of developing a device easy to use that can be an effective help to improve communication and inclusion of visually impaired population. The prototype works real-time processing the incoming video stream to the purpose of locating the presence of people and spotting known faces. Each event of interest produces a simple audio feedback to the user, allowing him or her to locate the presence of people before they start talking or highlighting known faces in noisy environments. So far the prototype has been quantitatively validated through a set of experiments carried out in lab; also a qualitative evaluation by a heterogeneous group of perspective users, users’ relatives, social workers, and experts gave us many positive comments and useful feedbacks for future work.

## I. INTRODUCTION

Visually impaired people continuously experience problems in integrating with a seeing society and difficulties due to a lack of visual control of the environment. Usually after years of training they learn how to conduct a reasonably independent life, assuming that no major variation intervene in their daily schedule. In less controlled circumstances they may have problems in finding their way in unknown places or to avoid physical barriers; it is more difficult for them to choose their clothes, to buy objects. An important and quite general difficulty in their daily life is in interacting with known and unknown people in unfamiliar environments. In the former case, they do not have clues on the presence of a known person in the environment, unless the other person sees them and sends them a vocal feedback; in the latter case, they may not know there is a person in their vicinity or they do not know whether this person is paying attention to them or not. This lack of visual contact greatly reduces the spontaneity of interpersonal contacts. These limitations are difficult to circumvent and often people learn to “do without” them, limiting their actions in order to prevent failures and frustrations.

The computer vision literature offers a vast selection of methods related to human faces, falling under the umbrella of face biometry that can be fruitfully exploited to address these issues. Specifically, *face detection* (i.e. the identification of the face of a person in the current scene) and *face recognition* (i.e. the identification of the identity of a person in the current scene). The main purpose of this work is to design,

develop and test a face biometry system intended for visually impaired users. The requirements coming from the specific application domain are many: first of all the devised solution should be easy to use since we cannot assume a technical attitude, second it should be low-cost and third it should quickly deliver information on the environment through audio messages. The PC-based prototype we developed so far is the first achievement of an ambitious project devoted to the design of a usable and accessible embedded system to be worn by the user.

The work we carry out is fueled by the experience we acquired in the last few years on developing real-time face detection methods [1], [2], face verification methods for access control [3], and designing usable and accessible tools for visually impaired users [4], [5], [6]. The main achievement of the work is on dealing with a specific and challenging application domain. With respect to previous work we inherit methods for real-time video processing, for face detection and verification in cluttered environments; also we find new applications to work that we previously carried out on developing easy-to-use systems for visually impaired.

Specific challenges that we are dealing with are:

- a moving camera whose position is controlled by a blind person and, as a consequence an increased amount of noise and a wider selection of illumination conditions;
- the need for a very simple training phase;
- the need to provide limited and essential vocal feedbacks and avoid audio spamming;
- a final application which is not as delicate as security or access control but it is addressed to users with very high expectations.

Some of these challenges have been addressed by choosing a specific hardware (for instance to deal with illumination issues), other by appropriate software solutions (a moving camera is not an issue since image analysis is done at a frame level), other required an *ad hoc* parameter tuning. In most cases the work we are carrying out is leading to an entire re-engineering of the technology previously developed, with a specific care to users’ needs.

The paper is organized as follows. Section 2 discusses related work on face detection and recognition and their application to the domain of visual impaired users. Section 3 describes the prototype developed so far, with particular reference to the face detection and recognition phases; Section 4 reports the experimental analysis carried out, while Section 5 summarizes the results of the activity of a working group

evaluating the quality of the work and the possible impact to perspective end users. Section 6 is left to a final discussion and an anticipation of future works.

## II. RELATED WORK

Face detection and recognition are among the most relevant issues in face biometry. Face detection is seen as a very delicate and complex preprocessing step to further analysis and a glimpse to such research area can be found in a dedicated web page, *Face Detection homepage*<sup>1</sup>, and in [7]. For face recognition, we refer the reader to the survey [8] or the material available on the *Face Recognition homepage*<sup>2</sup>. Face recognition methods based on intensity images may be roughly divided into global or holistic approaches and local or feature-based approaches. Eigenfaces [9] are with no doubt one of the most popular holistic approaches to face detection and recognition. In a training phase a relatively small set of data are used to estimate a subspace that permits to reduce the dimensionality of data with a quite small loss of information. The subspace is obtained with Principal Components Analysis (PCA). In [10] the eigenface method is extended to use a probabilistic similarity that models the *intra-personal* variations versus the *extra-personal* variations. The former are related to variations within the same individual, due to illumination, view-point, expression changes; the latter refer to the difference between an individual and another. An alternative approach of modeling *individual eigenspaces* each one describing a given individual have been proposed and shown effective for face authentication (see, for instance, [11]). The use of Fisher Discriminant Analysis as classifiers for face recognition is widely spread [12], [13], [8]. In [8] it is reported that according to [14] high frequency components, while they are not crucial for instance for the sex judgement task, may be important for face recognition, where finer details may help discriminating different identities. At the same time it has been shown that, especially when dealing with higher resolution images, global approaches are outperformed by component-based methods, that are more stable to local changes [15]. Feature-based approaches include methods for precise location of specific points (usually called *fiducial points*) [16], [17], [18] and methods based on templates, where each area of interest is described by approximately located patches. [19] compares the two approaches showing that the template based approach is simpler and better performing, while the approach based on precise features may lead to more compressed descriptions. Automatic feature extraction may be obtained extracting meaningful keypoints from the face image and describing them so to maximize the extra-class variance. We mention one of the first attempts to our knowledge to export the very popular SIFT descriptors to face recognition [20] and, the widely used Local Binary Patterns (LBP) [21] of which a number of variants and combinations are currently available.

Many technological solutions have been proposed for specific needs of visually impaired people, in particular for

wayfinding and more recently for automatic reading. Outdoor wayfinding is often based on the well established GPS technology (see, for instance, the projects [22], [23]). As for indoor navigation it is worth mentioning the Trinetra Project (CMU) based on barcode reading for object recognition [24]. Environment marking by means of RFID tags has been proposed [25], while the commercial product UltraCane5, based on a standard white cane, is equipped with a ultrasound system that estimates the distance of objects by computing the waves time-of-fly. The use of (passive) markers has been studied to increase the performance of computer vision approaches [26]. Systems based on computer vision have been proposed for their versatility as navigation aid in traffic intersections [27], [26] and sign reading [28]. Unfortunately, none of this computer vision research per se is yet practical for commercial use because of issues such as insufficient reliability and prohibitive computational complexity (which is especially problematic when using portable/wearable hardware). The only noticeable exception is portable OCR, such as the Abbyy OCR, available today on most cell phones. Among recent projects it is worth mentioning the iCare project<sup>3</sup> developing a set of digital tools for blind people, and in particular proposing automatic face recognition to address prosopagnosia also known as face blindness. In [29] a face recognition method to improve socialization skills of visually impaired population is proposed. To the best of our knowledge [29] is the only project addressing the same issue we deal with in this paper.

## III. THE PROTOTYPE

The developed prototype acquires and analyses a video stream real-time, detecting human faces in the scene, with particular reference to quasi-frontal faces approaching the camera/user. If more than one frontal face is detected the following procedure is repeated for each one of them. The face recognition module tries to associate a detected face with a dataset of models of people known to the user. If a similarity with a known person is highlighted, an audio feedback is returned to the user (e.g. "There is Augusto"). If the person is not recognized a feedback on the presence of an unknown person is sent ("there is a person" or "there is a person approaching").

Since it is important to limit the amount of information delivered to the user (to minimize audio spamming) the observations gathered over  $N$  video frames are merged before the actual feedback is sent. Thus, the recognition module analyses faces tracked over time and assigns an identity to current observations if a number of subsequent frames equals to  $N = 5$  returns a coherent feedback.

The system I/O is very simple:

- *Input* simply allows to switch on and off the detection and recognition modality.
- *Output* returns both audio and video feedback on the presence of people and their identity (if known).

<sup>1</sup><http://www.facedetection.com/>

<sup>2</sup><http://www.face-rec.org>

<sup>3</sup><http://cubic.asu.edu/projects/index.php>

In this section we give a brief overview of the prototype, starting from a description of hardware components, followed by details on face detection and face retrieval/recognition.

### A. The prototype hardware and software

While it is intended that the final goal is to implement a small size embedded system, the current prototype is based on a larger embedded architecture composed by three main elements:

- A PC fanless Celeron M 1Ghz, RAM 1GB DDR, HD 2,5" 160 GB, BTTV chip with 4 video channel.
- A compact Pixim ATM camera, Wide Dynamic Range, 504TVL, 2.9mm, 12VDC/24VAC.
- A speaker set for synthesizing audio messages.

The PC runs Linux Debian Etch and hosts a MySQL database storing records of known people and of observed events (people met while the system was running). The former are currently used for recognition, the latter are collected to be used in a semi-supervised setting to enrich the models of known people and possibly build automatically new models. These functionalities will be the core of future research.

### B. Face detection

First the video frame is segmented with a simple skin detection module, then potential skin regions are analysed with a face detection algorithm (see Fig. 1) that is tolerant to illumination variations. Our face detection module implements a cascade approach we recently presented in the literature [1], [2] and based on a feature selection plus classification pipeline: the feature selection procedure, applied to a high dimensional feature vector obtained from an over-complete set of rectangle features [30], allows us to automatically build from data an appropriately compact representation of the object of interest (faces in this case). Classification is implemented as a cascade of Support Vector Machine (SVM) classifiers [31] and allows for a fast detection on the image regions positive to the skin detection test.

Thanks to the data-driven nature of the whole pipeline, the same architecture is applied to train a eye and a nose classifier. The only difference here is in the choice of the specific training set. Eyes and nose detection are applied to the sole face region, allowing us to identify the most significant face images for subsequent recognition. Fig. 1 shows face detection and face feature detection results. Fig. 2 shows two faces detected over same video sequence: the frame on the left clearly shows all face features and therefore it is preferred during recognition.

### C. Face retrieval and face recognition

Once a human face is identified in the environment the face retrieval module is in charge of associating an identity to it. The working assumption useful for the case under analysis is that the number of people known to a single user is relatively small (a few tenths), but very many unknown people may be met by the user. Therefore, it is important we control the number of false positives of the system.



Fig. 2. Detecting face features within the face region allows us to discriminate unoccluded faces in a video (left) more appropriate for the following face retrieval.

At the same time a real-time feedback is an obvious requirement. Since a simple solution has to be preferred we design a two-layer architecture that can be summarized as follows:

- a first feedback by face retrieval on a set of  $N$  subsequent faces;
- a refinement of the answer by face validation.

The first phase aims at associating the most likely identity to the observed person; this impression is then verified by a face verification module. As for the second stage, we are currently implementing a simple SVM but we are in the process of adapting the face validation method proposed in [3] to the case under analysis. This will allow us to build specific representations and models for each known person, with a clear improvement in performance.

An important issue in maintaining a trade-off between precision and performance is in the choice of an appropriate initial data representation. We consider a local-based approach, that is, we describe the whole image as a concatenation of descriptions of image parts, starting off from the image region obtained from face detection and then we register the meaningful face area with the help of the eye region. We then adopt a rather standard feature-based representations based on the Local Binary Patterns (LBP) [21].

The choice of such representation is motivated by a comparative analysis with other local descriptions for face recognition, including variations of LBP such as Local Ternary Patterns (LTP) [32] and Histogram of Gradient (HoG) [33]. The performance of LBP is slightly superior, their computational cost smaller, and the description they produce is more compact. For all these reasons they have been preferred in current implementation. The similarity between images is based on a histogram intersection similarity measure [34] during the face retrieval phase. Although a histogram intersection kernel could be used in a SVM architecture [35], for computational reasons we implement a linear kernel.

More in details, given a *probe* image from the currently analysed video, we first compare it with all images in the *gallery* and rank the obtained results. We then analyse identities related the first  $N$  gallery images. If the number of images carrying a coherent identity  $I$  is above a given threshold  $\tau$  we associate the potential identity  $I$  to the probe and verify it by means of a binary SVM specific for  $I$  (as in a face authentication procedure [2]). Notice that we do not assume we have enough images to train an SVM classifier per each





Fig. 1. Face and face features detection in different illumination conditions: natural lighting on the top row, neon lights on the bottom row. The centre and right columns show binary maps output of skin detection, with (right) and without (centre) white balancing.

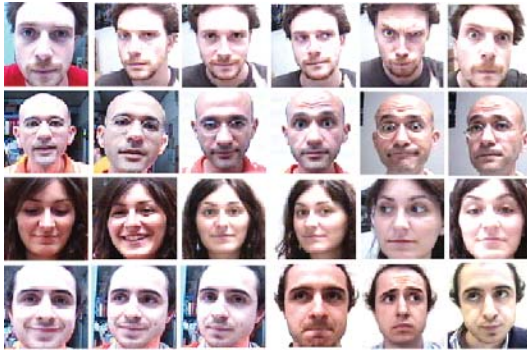


Fig. 3. Sample frames from the videos used in the experiments (the images have been automatically cropped with face detection). The environmental conditions differ, as well as facial expressions: during acquisitions people were ask to talk and move freely.

known person: in case a given identity is not associated a classifier we can still associate a potential identity to the probe, with a lower reliability.

#### IV. EXPERIMENTAL ANALYSIS

A set of laboratory experiments has been carried out to assess the appropriateness of the recognition pipeline for the application under analysis. The data used include about 200 video sequences all acquired by a still camera installed in 4 different environments (see Fig. 3), of which 100 contain a known person on the foreground, while the other 100 depict individuals unknown to the system. The number of known identities is 10. There is a n-to-n relationship between people and locations, meaning that a given person may appear in different environments.

To the purpose of the experiments the available data have been organized so that  $N$  image frames constitute the gallery, or model, of each known individual. We will refer to this set as the *training set*. Another set of data is used for validating the method and tuning the parameters (*validation set*). The remaining data are used as a *test set*.

Each data representation method is tuned so to choose the most meaningful set of parameters. All the reported experiments divide the image region under analysis in a  $7 \times 7$  grid;

method	false pos	false neg
LBP	0.15	0.21
HoG	0.22	0.29
LTP	0.42	0.39

TABLE I  
PERFORMANCE OF THE THREE DATA REPRESENTATIONS WITH RESPECT TO THEIR POWER IN DISCRIMINATING KNOWN PEOPLE FROM UNKNOWN PEOPLE.

LBP and LTP consider a pixel neighborhood of radius 3, while HoG a cell size of 3 and a block size of  $6 \times 6$  pixels. LTP is implemented without the normalization suggested by [32] since it does not appear to suit environment changes and abrupt illumination changes.

Fig. 4 shows a comparison among LBP, LTP and HoG, in terms of Receiver Operating Curves (ROC), built as follows: for each known identity we consider a set of negatives built as a merge of the positives of all the other identities and of images of unknown people. Then, we compute the percentage of false positives and hit rate. Finally, we average the obtained results over all the known identity. Each point of the curve is associated to a different value of the threshold  $\tau$ . The experiment, delivers an overall impression of the appropriateness of each classifier, and speaks in favor of LBP descriptions.

Fig. 5, instead, illustrates the confusion matrices obtained with the three methods. In this case it is worth noticing how LTP appear to be unappropriate for the task, since they not only propose a very low accuracy, but also a high degree of confusion among available classes. As a final experiment we explicitly refer to the discriminative power of the proposed method in classifying among known and unknown people. Table I reports the obtained results for the threshold  $\tau$  corresponding to a 5% false positive rate (over the validation set). Again, we may observe how LBP appear more appropriate. Fig. 6 shows a few successful recognition examples in different acquisition environments.

Current implementation of the method is based on a local description of the whole face area. An improvement, both from the point of view of efficiency and performance, would



Fig. 6. Sample frames showing successful recognitions.

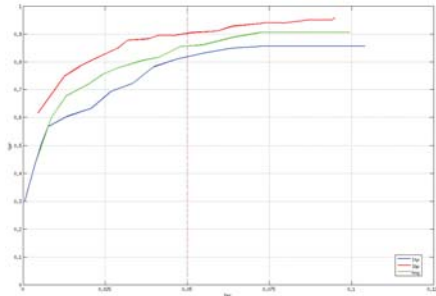


Fig. 4. Comparison of the overall retrieval performance for the three data representations considered: LBP, LTP, and HoG. The curve is built by varying the threshold  $\tau$ .

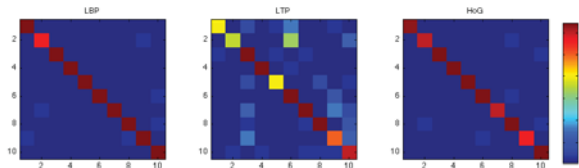


Fig. 5. Comparison of the confusion matrices for the three data representations considered: LBP, LTP, and HoG.

be to apply a feature selection phase to reduce the size of the initial data description. This approach has been previously explored by the authors for the case of face validation and showed a superior performance with respect to other well known dimensionality reduction methods, such as PCA (we remind the reader that PCA was adopted by [29] as the most appropriate method within a similar application to the one we are discussing). Experiments reported in [3], that we briefly report here for completeness, shown a superiority of local methods with respect to PCA: while PCA achieved an acceptance rate (AR) of 0.68 and a rejection rate (RR) of 0.7, a local method based on LBP allowed us to achieve  $AA = 0.87$  and  $RR = 0.94$ . A further improvement was achieved via feature selection; in this case  $AA = 0.93$  and  $RR = 0.97$ .

As for the robustness with respect to view-point changes, although the prototype was mainly devoted to analysing quasi-frontal faces, we experimentally observed that it is robust to angles of about 30 degrees. When a rotation is greater than this, face detection can be successful but face recognition is

not. Temporal coherence is important to keep track of overall result. Fig. 7 shows a typical behavior of the model under view-point variation.

The qualitative analysis reported in the next section has been carried out asking the users to try the system both with a still and a moving camera. From the technical point view it is maybe the case of stressing the fact that, since the face detection phase is based on the analysis of a single frame, there is no difference in processing a video from a still or a moving camera. The degradation of the video quality that would affect the obtained results appears when the camera has a considerable speed (something that with a hand held camera one would achieve by shaking the camera without control).

## V. THE ACTIVITY OF THE USERS GROUP

As the design and the development of the system prototype proceeds, a users group made of volunteers (both people with visual impairments or their relatives) and staff from Istituto Chiossone and Cooperativa Chiossone (Genova, Italy) are working hard to collect feedbacks from perspective users and to develop a procedure for prototype testing.

The first part of the activity, carried out between January and May 2010, consisted in interviewing perspective users about their interest on the technology under development. Chart 8 summarizes the profile of the interviewed 72 people, whose age is in the range [15 – 85] average 44,7, median 41,5.

The chart reported in Fig. 9 shows a summary of the answers given by the interviewed people. It is worth commenting the fact that while relatives often highlight issues related to the safety of the user (for instance, they would find interesting a face verification functionality with the purpose of checking declared identity of people), visual impaired users are more interested to the social potential of the technology under development. Moreover, the chart points out the needs of a technology able to detect people approaching the user with respect to the individuals that are present in the scene. By individual interviews, finally, someone pointed out the need to identify the moment in which in a pub the barman is looking at them waiting for their orders.

A qualitative evaluation of the prototype is being carried out by a user group coordinated by Istituto Chiossone staff. The system has been evaluated with respect to the effectiveness of detection and recognition, the feedback timing, and the robustness to view-point changes. The chart in Fig. 10 reports the

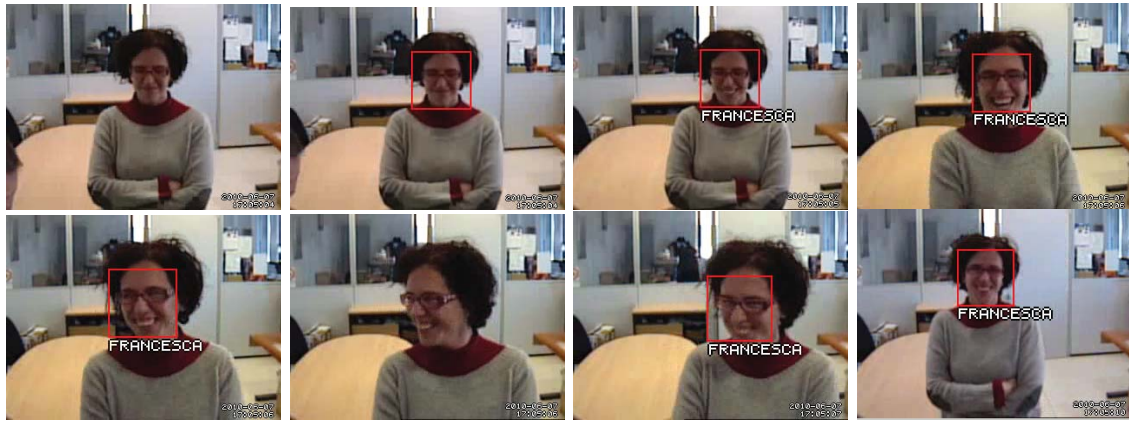


Fig. 7. Sample frames from a video sequence where the subject speaks freely, moving closer and farther the video camera, and rotating her head. Face detection fails when the head rotates more than 30 degrees approximately, but the temporal coherence applied to both detection and recognition allows us to keep track of the observation.

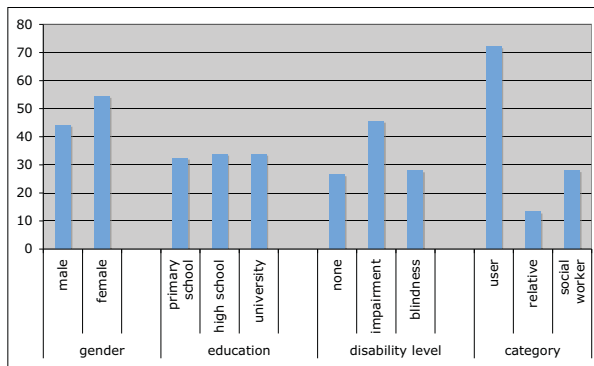


Fig. 8. Summary profile of the interviewed people (the chart reports percentages of the 72 people in the users' group)

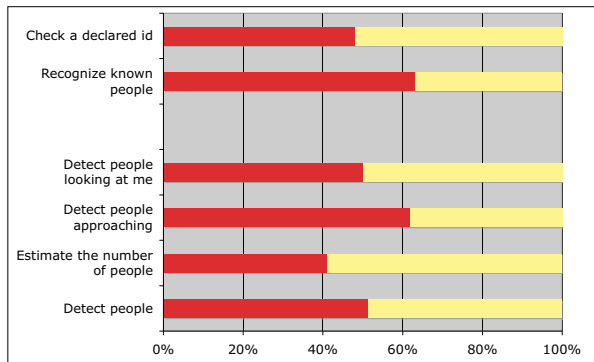


Fig. 9. Summary of the answers given by the interviewed people about their interest and their expectations on the project.

overall impression received so far — this working group also includes a small group of users expert on human and machine interface. The first group of questions have been addressed to all the interviewed people (visually impaired, relatives, social workers, technical experts and HMI experts), while the second group of questions has been addressed to people with no visual impairment, since they required a visual feedback. A few comments on the chart are in order. In general we received a very positive feedback on face detection, feedback speed and the absence of side effects when the camera was moving. As for face recognition we received very useful comments, but this functionality is still under development, as discussed in previous section. For what concerns the evaluation of difficult scene (we mean by that, very cluttered environments and the presence of many people — e.g., meetings) it has to be said that non technical users have very high expectations and tend to be disappointed by the system limitations. In this case the main problem was that the system was not calibrated to detect distant people. We received rather negative comments on the system feedbacks, but it is the case to point out that current I/O is very basic and certainly needs an improvement. As for the robustness to view-point changes, many people tried a profile posture, even if we explicitly said that the system is meant for quasi-frontal.

## VI. CONCLUSIONS AND FUTURE WORK

The paper presented a work in progress on a face recognition system developed for visually impaired users. The final goal of the project will be a dedicated embedded system to be worn by the user. Current prototype is instead a dedicated system of a medium size (easy to carry but impossible to wear) to be used to evaluate the system under different circumstances and seek users' opinions. Ongoing work is mainly devoted on extending the experimental dataset and on implementing clustering-based indexing methods to improve the efficiency of the face retrieval module as the gallery size grows. Also a feature selection method is being implemented for the face recognition module. From the users group stand point the main



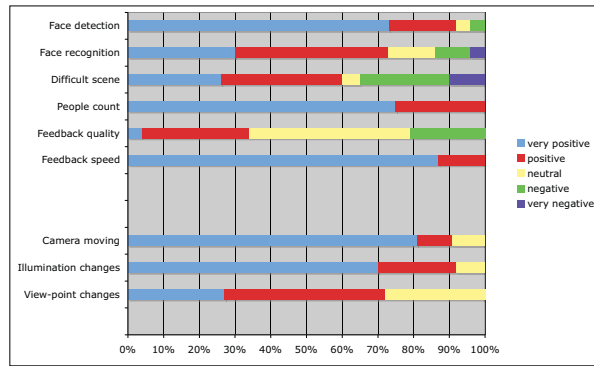


Fig. 10. Summary of the feedbacks received by the user group on the prototype developed so far.

effort is in delivering feedbacks concerning current prototype to the purpose of developing an embedded prototype.

The training phase during which the system learns to recognize a new person, is a crucial aspect of the work under development. Unlike [29] we aim at a simple procedure to gather the training data, ideally something that the user could do by himself, possibly with the help of a friend. Our model may start from a very small set of images of the person (say 10), that could be easily captured, and no requirement on the person pose or the scene illumination is made. In this case the first stage of recognition is activated automatically, while the second one is idle since there are not enough images to train the learning module for face verification. Further images can be added to the system and to the model among the ones that the system (correctly or erroneously) classifies as belonging to the person under analysis.

The prototype system is receiving many positive feedbacks by the users, since it delivers very basic but useful information concerning the surrounding environment without the need for a special training. Following this interest, a number of applications have been identified by the Istituto Chiossone staff working with visually impaired people. We mention a few of them:

- A small embedded module to be worn by the user (e.g., on his glasses) and delivering feedbacks online or on demand. This could be particularly important when the user enters a public place (for instance a cafe or a shop) and looks for some help, as a surrogate of the “eye touch”.
- A “talking” door bell, announcing the arrival of known people (so that the user does not have to enquire the new comer on his/her identity).
- A module running on the user PC that triggers a personalization of the desktop and the activation of additional functionalities and devices – this is especially important for computers in public areas (like a library or a multimedia room) with multiple users.

In the future we plan to investigate these application scenarios and to evaluate their usability. Special attentions will be devoted to the audio feedbacks returned to the users and also the input modalities (possibly given through vocal commands). They should be kept simple and reducing (as much as possible) audio spamming.

#### ACKNOWLEDGMENT

This work has been sponsored by Fondazione Carige (<http://www.fondazionecarige.it>) under the project *Riconoscersi e riconoscere: strumenti tecnologici di ausilio visivo (Recognize and self-recognize: technology as an aid of visually impaired users)*.

#### REFERENCES

- [1] A. Destrero, C. D. Mol, F. Odone, and A. Verri, “A sparsity-enforcing method for learning face features,” *IEEE Trans. on Image processing*, vol. 18, pp. 188–201, 2009.
- [2] —, “A regularized framework for feature selection in face detection and authentication,” *International Journal of Computer Vision*, 2009.
- [3] A. Destrero, A. Lovato, and F. Odone, “A prototype system for unconstrained face verification based on statistical learning,” in *Proceedings of the workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition - satellite event of ECCV*, 2008.
- [4] P. Boccacci, M. Ribaudo, and M. Mesiti, “A collaborative environment for the design of accessible educational objects,” in *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and International Conference on Intelligent Agent Technology*, 2008, pp. 207–210.
- [5] M. Mesiti, S. Valtolina, A. Bozza, M. Ribaudo, and S. Dini, “Accessibility and usability of a collaborative e-learning application,” in *Proceedings of the Second International Conference on Computer Supported Education*, 2010.
- [6] M. Mesiti, M. Ribaudo, S. Valtolina, R. Barricelli, P. Boccacci, and S. Dini, “Collaborative environments: Accessibility and usability for users with special needs,” in *Community-Built Database: Research and Development*, E. Pardede, Ed. Springer, In press.
- [7] M.-H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: a survey,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.
- [8] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, “Face recognition: A literature survey,” *ACM Computing Surveys*, pp. 399–458, 2003.
- [9] M. A. Turk and A. P. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [10] B. Moghaddam and A. Pentland, “Probabilistic visual learning for object representation,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 696–710, 1997.
- [11] X. Liu, T. Chen, and B. V. Kumar, “On modeling variations for face authentication,” *Pattern Recognition*, vol. 36, no. 2, pp. 313–328, 2003.
- [12] K. Etamad and R. Chellappa, “Discriminant analysis for recognition of human face images,” *Journal of the Optical Society of America A*, vol. 14, pp. 1724–1733, 1997.
- [13] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces versus fisherfaces,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 711–720, 1997.
- [14] J. Sergent, “Microgenesis of face perception,” *Aspects of face processing*, 1986.
- [15] B. Heisele, P. Ho, J. Wu, and T. Poggio, “Face recognition: component-based versus global approaches,” *Computer Vision and Image Understanding*, vol. 91, pp. 6–21, 2003.
- [16] A. Pentland, B. Moghaddam, and T. Starner, “Estimation of eye, eyebrow and nose features in videophone sequences,” in *International Workshop on Very Low Bitrate Video Coding (VLBV 98)*, 1998, p. 101104.
- [17] L. Wiskott, J. Fellous, N. Kuiger, and C. von der Malsburg, “Face recognition by elastic bunch graph matching,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 775–779, 1997.
- [18] R. L. S. Arca, P. Campadelli, “A face recognition system based on automatically determined facial fiducial points,” *Pattern recognition*, vol. 39, no. 3, pp. 432–443, 2006.

- [19] R. Brunelli and T. Poggio, "Face recognition: features versus templates," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042–1052, 1993.
- [20] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli, "On the use of sift features for face authentication," in *Proc. of IEEE Int Workshop on Biometrics, in association with CVPR06*, 2006, p. 35ff.
- [21] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Trans. on PAMI*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [22] L. Ran, A. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," in *Proc. of the 2nd IEEE Pervasive Computing Conference*, 2004.
- [23] B. N. Walker and J. Lindsay, "Navigation performance with a virtual auditory display: Effects of beacon sound, capture radius, and practice," *Human Factors*, vol. 48, no. 2, pp. 265–278, 2006.
- [24] P. Narasimhan, "Trinetra: Assistive technologies for grocery shopping for the blind," in *IEEE-BAIS Symposium on Research in Assistive Technologies*, 2007.
- [25] H. Lee, M.-S. Tsai, M.-C. Wang, and J. Y. Hsu, "icane - a partner for the visually impaired," in *2nd International Symposium on Ubiquitous Intelligence and Smart Worlds (UISW 2005), Embedded and Ubiquitous Computing Workshops (EUS)*, 2005.
- [26] J. Coughlan and R. Manduchi, "Color targets: Fiducials to help visually impaired people find their way by camera phone," *Eurasip Journ. on Image and Video Processing, Special Issue on Image and Video Processing for Disability*, 2007.
- [27] M. S. Uddin and T. Shioyama, "Bipolarity- and projective invariant-based zebra-crossing detection for the visually impaired," in *Proceedings of the 1st IEEE Workshop on Computer Vision Applications for the Visually Impaired*, 2005.
- [28] P. Silapachote, J. Weinman, A. Hanson, R. Weiss, and M. A. Mattar, "Automatic sign detection and recognition in natural scenes," in *Proc. of the 1st IEEE Work on Computer Vision App. for the Visually Impaired (CVAVI 05)*, 2005.
- [29] S. Krishna, G. Little, J. Black, and S. Panchanathan, "a wearable face recognition system for individuals with visual impariments," in *ASSETS'05*, 2005.
- [30] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal on Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [31] V. N. Vapnik, *Statistical Learning Theory*. Wiley, 1998.
- [32] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions." in *AMFG*, 2007, pp. 168–182.
- [33] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE CVPR*, 2005, pp. 886–893.
- [34] M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, vol. 7, pp. 11–32, 1991.
- [35] F. Odone, A. Barla, and A. Verri, "Building kernels from binary strings for image matching," *Image Processing, IEEE Transaction on*, 2005.