# Image Processing Issues in a Social Assistive System for the Blind

Margherita Bonetto, Sergio Carrato, Gianfranco Fenu, Eric Medvet,
Enzo Mumolo, Felice Andrea Pellegrino, Giovanni Ramponi
DIA, University of Trieste, Italy
Email: bonetto.margherita@gmail.com, {carrato, fenu, emedvet, mumolo, fapellegrino, ramponi}@units.it

***Abstract*—We systematically analyse the design of the low-level vision components of a real-time system able to help a blind person in his/her social interactions. We focus on the acquisition and processing of the video sequences that are acquired by a wearable sensor (a smartphone camera or a Webcam) for the detection of faces in the scene. We review some classical and some very recent techniques that seem appropriate to the requirements of our goal.**

## I. Introduction

The loss of vision poses several challenges to impaired people, since they are children and adolescents [1]. An extremely important one is to be able to build rich and natural interpersonal relationships; sight loss is indeed reported to inhibit social interaction [2], [3]. The blind person can feel lonely, isolated and then depressed; the impaired person can feel that he/she has nothing to offer to other people.

A research project has been recently funded by the University of Trieste and a private donation, which aims to devise user-friendly vision-based techniques that may assist the social interaction of a person affected by a very severe visual impairment or a total blindness. The first phase of the project requires to build a system view of the information acquisition and processing chain, and to determine the properties of its various components: from the acquisition of video data, to data pre-processing and analysis, the extraction of non-verbal information that the blind person cannot perceive, its summarization and communication to the user through an audio channel. In this paper several of these components are briefly mentioned, while the attention is focused on a subset of tools, whose performance are fundamental to achieve the project objectives: the processing of the video sequences acquired by a wearable sensor with the purpose of detecting faces in the scene. We review some classical and some very recent techniques in the field, and revisit them according to the requirements of our goal; the originality of this contribution is related to the specific application we refer to. Section II discusses the characteristics of *egocentric* (aka First Person Vision) videos as they are acquired by a blind person; other peculiarities of the video acquisition are examined in Section III. Preprocessing issues are dealt with in Section IV, which are related to the face detection techniques discussed next in Section V.

## II. First person vision methods

"It sees what the user sees and looks where the user is looking" [4] is a well known definition of what is, in general, the main characteristic feature of a First Person Vision (FPV) system. Indeed, according to the aims of our proposal, the claim should be modified as: *it sees what the user would see and looks where the user would be looking*. In fact, we are not interested in the users' behaviour and intentions, but on supporting them and enhance their social interactions. Paying attention to this goal, we report the main features (pros and cons) of FPV systems, the already available configurations and algorithms, challenges and desiderata related to the proposal context. For a very recent and complete review of FPV in general, the reader may refer to [5].

### *A. Characterisation and challenges*

FPV videos have a set of characteristic features that differentiate egocentric videos from classical videos acquired by a third-person point of view. Of course, the camera alone is not sufficient to classify a video as FPV video: a video acquired using a recent smartphone, kept steady when shooting, is not an egocentric video. Moreover, even the semantic content alone does not permit to classify a video as FPV video: simulating the point-of-view of a character in a film or TV-drama, guaranteeing professional video quality, is not a true FPV video. Indeed, a moving camera is a requisite to produce a FPV video, but there are some other important low-level features characterising FPV videos. As pointed out, among others, in [6], [7], [8], low-level peculiarities of an egocentric video are:

- the video is recorded as a long sequence, without cutting it into shots (as usually done in standard edited videos);
- the camera is moving, since it follows the movements of the person: these movements lead to motion blur effects in the video;
- moreover, camera movements may also lead to rapid changes and strong variations in illumination, for example when the person leaves a room or he/she faces a window.

Taking into account the proposed scenario, FPV videos provide a wide range of challenges, such as restoring motion blur, compensating for strong variations in illumination, and performing video summarization. The first two can be coped with using suitable image preprocessing, as we shall show below. For what concerns video summarization, it has to be noted that in the proposed scenario FPV videos are data streams that do not need to be stored, but to be analyzed quickly. However, the tools developed in the video summarization literature may be exploited for detecting the relevant (w.r.t. some criteria) changes that need to be communicated to the user.


Image Processing

### B. Goals

Among the various issues related to FPV videos, which are analysed in the literature (see [5]), the following topics are relevant for the present analysis:

*1) Localization – Where is the user right now?:* There is not yet a general FPV video based localization algorithm applicable in every outdoor and indoor environment, and a review of the state-of-the-art of localization techniques, based on FPV videos, is outside the scope of the present paper. Here it is worth noting that the most common approach for localizing a person, by means of FPV videos, is based on similarity between the current observed image and the images in a set of previously acquired images (see [4] and the references cited therein). Indeed, without image preprocessing the undesirable features of FPV videos may affect the performance of the algorithm evaluating the similarity between the current image and the image data-set. Recently, in [9] the authors proposed an evolution of such an approach, using not only image similiarity but also sensor data from the user's device IMU[1] with the aim to obtain a more efficient matching and to minimize false matches. Once the user position has been estimated by image matching and sensor fusion, additional information may be obtained from the meta-data associated to the best matching image, such as directions, accessibility information, etc. Such pieces of information may be helpful to the user in the proposed scenario.

*2) Scene structure – Is there anybody out there?:* Location of people in the scene and eventually their identity are essential aims of our proposal. Real-time techniques for face detection and recognition are needed to implement such capabilities, but nowadays this is not *per se* an issue, even on smartphones (see for example [10], [11], [12]).

## III. Video acquisition

The challenges of FPV that we indicated in the previous section are further complicated by the fact that the particular users we are considering cannot exploit a visual feedback to control the quality and the contents of the acquired video. By the way, this also means that the acquired data substantially differ from those normally used in the training of the algorithms used for the detection and recognition of faces; the distance and direction of view (and consequently the apparent pose of people) and possible poor illumination conditions (e.g., backlights) are two specific issues. It may thus be necessary to resort to techniques of Visual Domain Adaptation (VDA) [13].

The main sensor device in every FPV system is a wearable camera [5], [4]. The position and the characteristics of the video camera need to be examined with care. Some researches [14], [15] describe models that depict how wearable cameras performance is affected by viewing angle and wearer motion. It is calculated that head, shoulders, arms, outer legs and even feet have the largest angle of views, followed by the chest and waist. When the user walks, the human face experiences the least motion, followed by the head, chest and stomach, then the upper forearms followed by the upper legs, and finally the extremities. However, for our application the mentioned models need to include in the angle-of-view calculations the condition that the faces to be detected are in most cases located in a limited height range. Moreover, a person affected by blindness often develops some behavioural mannerism ("blindism" [16]); for instance body rocking, clapping, head shaking, eye pressing and poking, etc. From a technical viewpoint, mannerisms drastically change the classification above about the camera location, e.g., making the head a totally unsuitable camera site. A camera attached to a pair of eyeglasses [4], [17], [18] is useful in the case of visually impaired people (the same target as in the BeMyEyes project [19]), but definitely not for many blind people. Another important aspect is the social acceptability of the camera: according to the crowdsourcing evaluation performed in [15], egocentric cameras on the head or face are less acceptable than cameras across the chest or on the ear. A recently proposed solution is to place the visual sensor on a cane [20], with the advantage of exploiting a tool that the user is already familiar with. Wearable cameras however permit to get rid of the cane at least in some environment contexts. In fact, to cope with mannerisms, the most effective positions for a wearable camera are shoulder (as in [10]) or, as suggested in [8], [7], the upper portion of the chest, e.g., in the lapel area. As stated above, in the latter situation the faces of very close people may be out of the camera field of vision, whereas the viewpoint obtained from a camera in the shoulder position is sufficiently close to the natural viewpoint of a seeing person interacting with other people.

With respect to the optic components, in [8], [7] a camera with a wide-angle lens has been proposed as wearable device for remotely monitoring patients with dementia diseases. Motivation of such a camera was the capability of capturing most of the scene facing the user, which might be a requirement also in the proposed scenario. However, a wide-angle lens is responsible of optical distortion in the acquired videos (the so called barrel distortion, briefly recalled in Section IV-D) and such a distortion has to be compensated, before starting to detect any kind of object or faces in the video frames.

## IV. Preprocessing

It is reasonable to expect that suitable image preprocessing techniques can lead to an improvement in system performance, in particular in case of poor image quality. For what concerns face detection, for example, algorithms based on the location of areas having different gray levels, as the one by Viola and Jones [21], suffer from poor illumination condition and consequent low dynamics of the useful part of the signal and could benefit from a suitable dynamic correction/expansion; similarly, algorithms based on the location of facial landmarks perform poorly in case of noisy pixels that are likely present in low light condition, and they would yield better performance with a suitable edge preserving smoothing preprocessing.

Of course, issues related to the computational complexity have to be taken into account. It has to be noted that, while the bandwidth of the communication channel towards the user is very limited (information should be delivered to the blind person, say, once per second at most), it could be useful to process the incoming video at a higher frame rate, in order to also exploit, e.g., temporal correlation between temporally close frames. Probably a true real-time operation, i.e., 25 or 30

[1] All today's smartphones have an Inertial Measurement Unit (IMU) on board.

fps, is not necessary for our purposes, and a frame rate around a few frames per second is a reasonable requirement.

For the case we are studying, preprocessing could be mainly considered with reference to the correction of poor lighting condition, global motion deblurring, global motion compensation, and correction of barrel distortion in wide angle lenses.

### A. *Dynamic range of the luminance*

Issues related to the dynamic range of the luminance are typically present when both very light and very dark areas simultaneously appear in the scene. This may easily happen both in indoor environments [22], e.g., if there is a window showing the outside world, and in outdoor ones, e.g., for backlit objects and people in a sunny day. High Dynamic Range (HDR) is not a problem *per se*, as indeed HDR images have more information than their low dynamic range (LDR) counterparts, but sensors in consumer electronic devices are normally not able to capture HDR images, so that a significant amount of information is lost in the acquisition process. The trick which is commonly used to get HDR images in these devices, i.e., to suitably merge information from multiple LDR images obtained with different time exposure settings, may lead to unacceptably large computational load and power consumption if large motion of the subject and/or of the sensor is to be compensated [23]. In case HDR images are actually available, these can yield to better results with respect to the use of LDR ones [22]. In the case of LDR images, dynamic range correction algorithms can be used to tackle the problem. Some approaches rely on the Retinex approach [24], and try to separate the illumination component (which we are not interested into) from the reflectance one. In [25] a novel processing method using contrast-limited adaptive histogram equalization (CLAHE) is applied to adjust the illumination in order to improve the naturalness of the image. Alternatively, in [26] the variations in the illumination components are reduced by working in the DCT domain, while the color constancy algorithm is used in [27] for the same purpose.

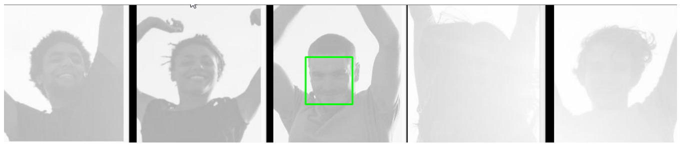
Improvements can also be expected if colour information is used. For example, in [28], morphological operators are applied to the color components, in the HSV space, to improve the performance of a full-body detection algorithm.

A simple experiment shows the effects of preprocessing operators applied to images with poor illumination condition. Figure 1a shows some portions of a photograph of a backlit scene. The standard Viola-Jones [21] face detection algorithm is not able to detect any face in this image. Applying a simple gamma correction to the data, one face is detected (Figure 1b); if both a suitable global illumination normalization and a gamma correction are applied, two faces are detected (Figure 1c). The same type of preprocessing has positive effects also on other face detection operators.

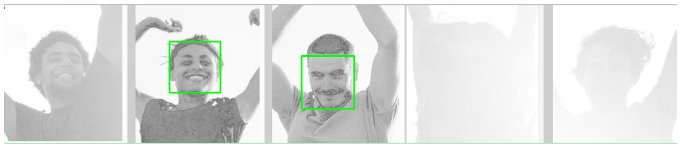
### B. *Global motion deblurring*

Deblurring is useful in case of low light level, when the consequent long exposure time can lead to motion blur if the sensor is subject to translation and, even more importantly, rotation; this blur can negatively affect the performance of the following algorithms, e.g., face detection. Both translation and rotation along an axis belonging to the sensor plane result in a panning effect, so that all the pixels in the image are subject to the same point spread function (PSF) and a conventional space invariant deconvolution can be applied; in this case, it is possible either to use blind deconvolution algorithms, such as [29], or to estimate the motion and then to follow approaches, such as Lucy-Richardson [30], [31], which require the knowledge of the PSF. Rotation along the axis orthogonal to the sensor plane, in turn, yields a rotational optical flow, which implies the use of space variant image deconvolution operators. Even in this case some of these, e.g., [32], [33], follow a blind approach, while others, such as [34], [35], require a prior estimation of the PSF in different areas of the image; in any case, computational complexity requirements could limit the applicability of these techniques.

(a) Original image



(b) Viola-Jones output after gamma correction



(c) Viola-Jones output after illumination normalization and gamma correction

Fig. 1: Impact of illumination on face detection.

### C. *Global motion compensation*

Motion compensation can be fruitfully applied if the subsequent operators also exploit temporal information or are sensitive to rotation, due to the fact that sensor motion, as already mentioned, can lead to a panning effect or a rotational optical flow. It is useful, for example, to reduce the number of false negatives and false positives of a face detection algorithm by analysing the correspondence of the detector output in the same area in temporally adjacent frames. Moreover, algorithms such as Viola-Jones [21] are severely affected by in-plane rotations above $\pm 15°$, so that at least a coarse rotation correction is needed. Conventional motion compensation algorithms such as [36] can be applied, taking into account that in this case a very accurate compensation is probably not necessary: for our purposes the subsequent algorithms are normally rather robust to small translations and rotations, so that computational complexity should not be an issue in this case.

### D. *Lens distortions*

As already mentioned, barrel distortion is expected when using wide angle lenses. Indeed, it does make sense to use


Image Processing

wide angle lenses in this kind of application, where the user is not able to verify the correct orientation of the camera; the large number of pixels in modern low-cost sensors should provide enough resolution also for details (e.g., the face of a person) which are rather far from the sensor. With reference e.g., to face detection, the barrel distortion is not a problem for far faces, as they occupy a limited area of the image, but it becomes an issue for close-up ones, if the subsequent face detection algorithm is sensitive to geometrical distortions. Fortunately, the lens distortion is normally known, given the make and model of the camera, or can be easily estimated by taking a picture of a suitable test image, and the corresponding correction can be easily applied in real time, using e.g., the algorithm proposed in [37].

## V. Face detection with poor image quality

A wealth of research effort has been devoted in the last decade to improve the effectiveness of face detection. Significant improvements have been obtained in particular in those conditions where image quality is not optimal: the so-called *in-the-wild* images exhibit great variations in pose, illumination, and occlusion of the faces and their abundant availability—enabled by the web and Online Social Networks—allowed researchers to develop robust face detectors.

In this section, we focus on those approaches which purposely addressed those image quality issues which are specific of the scenario considered in this paper, namely bad lighting condition, motion blur, and image distortion, as already mentioned in Section IV. Besides face detection, we include some approaches originally proposed for face recognition, that could be adapted for the purpose of detection. The last two paragraphs of the section are devoted to techniques that are robust to variations in pose of the face to be detected. For a wider analysis, we refer the reader to [38], which surveys the recent advances in real-world face detection techniques and the main databases used for the evaluation of face detection algorithms.

Algorithms exist which are *inherently* robust to illuminance disturbances. For example, semi-local structure patterns (SLSP) are proposed in [39] to get robust face detection, while in [40] Local Gradient Patterns (LGP) are proposed as an alternative to Local Binary Patterns (LBP), which are sensitive to local light intensity variations. A similar approach is followed in [41], where Boosted Local Binary (BLB) features are proposed which are based on both the gray level image and the gradient image.

Robustness to bad lighting conditions has been the focus of several works on face detection and recognition, the latter usually including the former as a premising step. In [42], a systematic analysis of several preprocessing methods aimed at improving face recognition is proposed. The authors assess 14 different illumination preprocessing methods coupled with 6 face recognition algorithms, using 4 public available datasets: their findings suggest that some holistic illumination approaches improves the performance of face recognition. Interestingly, the focus of [42] is on preprocessing methods which have to be applied after face detection and before face recognition—it follows that their role in our application has to be further investigated. However, the authors of the cited paper suggest that illumination preprocessing on skin areas detected using skin detection techniques can be a complementary approach w.r.t. the ones considered in their paper. Indeed, modelling human skin has been often proposed as an option to build illumination-robust face detectors. In [43] a skin model in the YCbCr color space is proposed which is stated to be robust to luminance variations and noise. In [44], a fusion strategy is shown which builds on an enhancement-based skin color segmentation approach: the authors experimentally show that the strategy can improve the performance of face detection and is robust against complex background, and against ethnicity- and lighting-related variations. A system which involves skin and face detection under rapidly changing illumination conditions is presented in [45]: the face detection step allows the system to adjust color correction, which is a premise to improve the accuracy of skin detection.

Motion blur is another issue which can affect images acquired in the scenario considered in this paper. As discussed in Section IV-B, deblurring can be performed leveraging the knowledge of the PSF: some studies have been conducted concerning the estimation of the PSF of face images. The authors of [46] propose a method which exploits a training set of blurred face images—for which the PSF is known—based on a feature space where blurred faces degraded by the same PSF are similar to one another. They evaluate their method on a large face database artificially degraded by focus or motion blur and show that it substantially improves the recognition performance. A similar approach is presented in [47], where the system uses a set of blurred face images where face edges are manually marked. In both cases ([46], [47]), however, deblurring algorithms act on a local image region which contains only the face: we hence argue that the applicability of those results to our scenario is not straightforward. A quite different approach for coping with blurred images is shown in [46]: instead of deblurring the image and then performing face detection/recognition, the authors develop a blur-robust descriptor. Under some assumptions concerning noise and the blur kernel, an image and its blurred version are equal in terms of the proposed descriptors: the recognition of faces is then performed directly on the blurred images using these descriptors.

Less research has been done on face detection methods which are robust to image distortion. A recent study [48] focuses on face detection in omnidirectional images, an acquisition sensor which might be of interest in our scenario. The main authors' finding concerns the use of proper descriptors, instead of intermediary representation of the image. In particular, they propose the use of adapted Haar-like features which are shown to be both descriptive and discriminant.

Other works systematically investigate some respects of image quality and their impact on face detection. In [49], the authors assess the effective degradation of a state-of-the-art face detector when human-perceived image quality is reduced by means of additive white Gaussian noise, Gaussian blur or JPEG compression: moreover, they augment their study with a Distorted Face Database and they propose a new face detector which is robust to the considered image quality degradations. A different point of view is instead taken in [50]: the authors experimentally show that the video quality metrics which are designed for human perception are not suitable for measuring the quality as perceived by face analysis algorithms, namely detection, recognition, and tracking. Hence, they propose two alternative metrics based on blockiness and mutual information.

Robustness to pose has always been an issue for the most widespread face detection methods: *multiview* face detection remains a challenging task, as the appearance under various pose, illumination and expression conditions change dramatically. Many datasets have been developed for testing multiview face detection algorithms; among them, the Face Detection Data Set and Benchmark (FDDB) [51] is used by many authors as a benchmark. The FDDB dataset includes 2845 images with a total of 5171 faces, both grayscale and color images. The faces are acquired considering a wide range of difficulties including occlusions, difficult poses, and low resolution and out-of-focus faces. Classic strategies, including [21] and [52], divide face images into multiple categories, e.g., frontal, half profile, profile, etc, and train different classifiers on these categories. However, the performance of classical face detection algorithms is still unsatisfactory, as shown in [51]. In the past few years deep neural networks (DNN) have attracted much interest due to their exceptional performance in multi-object class detection [53]; thus, DNN have been investigated for face detection [54].

Many authors use Convolutional Neural Network (CNN) face detection. CNN are very similar to ordinary Neural Networks [55]. The inputs of a CNN are images and its layers have neurons arranged in 3 dimensions: width, height, depth. Each layer transforms the 3D input volume to a 3D output volume of neuron activations. Compared with the previous hand-crafted features, CNN can automatically learn features to capture complex visual variations by leveraging a large amount of training data and can be easily parallelized on GPU cores for acceleration. Considering the relatively high computational expense of the CNNs, exhaustively scanning the full image in multiple scales with a deep CNN is not a practical solution. To achieve fast face detection, Li et al. propose in [56] a CNN cascade, which rejects negative instances quickly in the early, low resolution stages and carefully verify the detections in the later, high-resolution stages. Li et al. experimentally show that their approach outperform the state-of-the-art methods in face detection. In particular, they achieved about 85% accuracy with 100 false positives on the FDDB dataset. Farfade et al. [57] propose a face detection method based on deep learning, called Deep Dense Face Detector, which achieves about 82% accuracy on FDDB.

## VI. Conclusions

Some image processing tools have been analyzed, suitable for the low-level vision components of an assistive system for the blind. Video sequences that are acquired by a wearable sensor were addressed, revisiting the most recent techniques according to the requirements of our goal. We think this study can provide a guideline for the system-level design of the assistive equipment. Further study will be devoted to a similar analysis about computer vision tools to be deployed for more advanced tasks, such as face recognition and the recognition of facial expressions, posture, gesture, inter-personal interactions.

## Acknowledgment

This work has been supported by the University of Trieste - Finanziamento di Ateneo per progetti di ricerca scientifica - FRA 2014, and by a private donation in memory of Angelo Soranzo (1939-2012).

## References

[1] S. Kef, J. Hox, and H. Habekothé, "Social networks of visually impaired and blind adolescents. structure and effect on well-being," *Social Networks*, vol. 22, no. 1, pp. 73 – 91, 2000.

[2] I. Bruce, J. Harrow, and P. Obolenskaya, "Blind and partially sighted people's perceptions of their inclusion by family and friends," *British Journal of Visual Impairment*, vol. 25, no. 1, pp. 68–85, 2007.

[3] D. Bolt, "From blindness to visual impairment: Terminological typology and the social model of disability," *Disability & Society*, vol. 20, no. 5, pp. 539–552, 2005.

[4] T. Kanade and M. Hebert, "First-person vision," *Proceedings of the IEEE*, vol. 100, no. 8, pp. 2442–2453, Aug 2012.

[5] A. Betancourt, P. Morerio, C. Regazzoni, and M. Rauterberg, "The evolution of first person vision methods: A survey," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.

[6] C. Tan, H. Goh, V. Chandrasekhar, L. Li, and J.-H. Lim, "Understanding the nature of first-person videos: Characterization and classification using low-level features," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on*, June 2014, pp. 549–556.

[7] S. Karaman, J. Benois-Pineau, R. Mégret, J. Pinquier, Y. Gaestel, and J.-F. Dartigues, "Activities of daily living indexing by hierarchical hmm for dementia diagnostics," in *Content-Based Multimedia Indexing (CBMI), 2011 9th International Workshop on*, June 2011, pp. 79–84.

[8] S. Karaman, J. Benois-Pineau, R. Mégret, V. Dovgalecs, J.-F. Dartigues, and Y. Ga estel, "Human daily activities indexing in videos from wearable cameras for monitoring of patients with dementia diseases," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, Aug 2010, pp. 4113–4116.

[9] V. Bettadapura, I. Essa, and C. Pantofaru, "Egocentric field-of-view localization using first-person point-of-view devices," in *Winter Conference on Applications of Computer Vision (WACV 2015)*. IEEE, 2015.

[10] S. Chaudhry and R. Chandra, "Design of a mobile face recognition system for visually impaired persons," *CoRR*, vol. abs/1502.00756, 2015. [Online]. Available: http://arxiv.org/abs/1502.00756

[11] K. Kramer, D. Hedin, and D. Rolkosky, "Smartphone based face recognition tool for the blind," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, Aug 2010, pp. 4538–4541.

[12] Y.-C. Wang and K.-T. Cheng, "Energy-optimized mapping of application to smartphone platform – a case study of mobile face recognition," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, June 2011, pp. 84–89.

[13] V. Patel, R. Gopalan, R. Li, and R. Chellappa, "Visual domain adaptation: A survey of recent advances," *Signal Processing Magazine, IEEE*, vol. 32, no. 3, pp. 53–69, May 2015.

[14] W. Mayol-Cuevas, B. Tordoff, and D. Murray, "On the choice and placement of wearable vision sensors," *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, vol. 39, no. 2, pp. 414–425, March 2009.

[15] D. S. Hayden, "Wearable-assisted social interaction as assistive technology for the blind," Ph.D. dissertation, Massachusetts Institute of Technology, 2014.

[16] E. Fazzi, J. Lanners, S. Danova, O. Ferrarri-Ginevra, C. Gheza, A. Luparia, U. Balottin, and G. Lanzi, "Stereotyped behaviours in blind children," *Brain and Development*, vol. 21, no. 8, pp. 522–528, 1999.

[17] A. R. Doherty, S. E. Hodges, A. C. King, A. F. Smeaton, E. Berry, C. J. Moulin, S. Lindley, P. Kelly, and C. Foster, "Wearable cameras in health: The state of the art and future possibilities," *American Journal of Preventive Medicine*,

vol. 44, no. 3, pp. 320 – 323, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0749379712008665

[18] X. Wang, X. Zhao, V. Prakash, W. Shi, and O. Gnawali, "Computerized-eyewear based face recognition system for improving social lives of prosopagnosics," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*, May 2013, pp. 77–80.

[19] Be My Eyes. web site. http://www.bemyeyes,org/.

[20] Birmingham City University, UK. A pioneering facial recognition cane for blind people. http://www.bcu.ac.uk/news-events/news/a-pioneering-facial-recognition-cane-for-the-blind.

[21] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[22] P. Agrafiotis, E. K. Stathopoulou, A. Georgopoulos, and A. Doulamis, "HDR imaging for enchancing people detection and tracking in indoor environments," in *Proc. VISAPP2015*, 2015.

[23] T. Jinno and M. Okuda, "Multiple exposure fusion for high dynamic range image acquisition," *Image Processing, IEEE Transactions on*, vol. 21, no. 1, pp. 358–365, 2012.

[24] E. Land, "The retinex," *American Scientist*, vol. 52, no. 2, pp. 247–264, 1964.

[25] X. Fu, Y. Sun, M. LiWang, Y. Huang, X.-P. Zhang, and X. Ding, "A novel retinex based approach for image enhancement with illumination adjustment," in *2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, 2014.

[26] P. Goel and S. Agarwal, "An illumination invariant robust and fast face detection, feature extraction based face recognition system," in *Proc. 2012 Third International Conference on Computer and Communication Technology*, 2012.

[27] S.Ashok and K.K.Thyaghrajan, "Facial expression recognition with auto-illumination correction," in *Proc. 2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, 2013.

[28] H.-W. Chen and M. McGurr, "Improved color and intensity patch segmentation for human full-body and body-parts detection and tracking," in *Proc. 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2014.

[29] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, pp. 43–64, May 1996.

[30] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical Journal*, vol. 70, no. 6, pp. 745–754, June 1974.

[31] W. H. Richardson, "Bayesian-based iterative method of image restoration," *Journal of the Optical Society of America*, vol. 62, no. 1, pp. 55–59, Jan. 1972.

[32] S. Harmeling, M. Hirsch, and B. Schölkopf, "Space-variant single-image blind deconvolution for removing camera shake," in *NIPS 2010*, Vancouver, Canada, Dec 2010.

[33] M. Hirsch, S. Sra, and B. Schölkopf, "Efficient filter flow for space-variant multiframe blind deconvolution," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2010.

[34] M. L. Cobb, P. L. Hertz, R. O. Whaley, and E. A. Hoffman, "Space-variant point-spread-function deconvolution of Hubble imagery using the Connection Machine," in *Proc. SPIE 2029, 202*, 1993.

[35] S. J. Weddell and R. Y. Webb, "The restoration of extended astronomical images using the spatially-variant point spread function," in *23rd International Conference Image and Vision Computing New Zealand*, (New Zealand), 2008.

[36] F. Dufaux and J. Konrad, "Efficient, robust, and fast global motion estimation for video coding," *IEEE Trans. on Image Processingl*, vol. 9, no. 3, pp. 497–501, March 2000.

[37] Y. Altunbasak, R. M. Mersereau, and A. J. Patti, "A fast parametric motion estimation algorithm with illumination and lens distortion correction,," *Image Processing, IEEE Transactions on*, vol. 12, no. 4, p. 395, April 2003.

[38] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: past, present and future," *Computer Vision and Image Understanding*, 2015.

[39] K. Jeong, J. Choi, and G.-J. Jang, "Semi-local structure patterns for robust face detection," *Signal Processing Letters, IEEE*, vol. 22, no. 9, pp. 1400–1403, Sept 2015.

[40] B. Jun, I. Choi, and D. Kim, "Local transform features and hybridization for accurate face and human detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 6, pp. 1423–1436, 2013.

[41] H. Ren and Z.-N. Li, "Boosted local binaries for object detection," in *Proc. 2014 IEEE International Conference on Multimedia and Expo (ICME)*, 2014.

[42] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1691–1699, 2013.

[43] S. L. Phung, A. Bouzerdoum, and D. Chai, "A novel skin color model in ycbcr color space and its application to human face detection," in *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1. IEEE, 2002, pp. I–289.

[44] R. Mishra and R. Subban, "Face detection for video summary using enhancement-based fusion strategy under varying illumination conditions," in *Science Engineering and Management Research (ICSEMR), 2014 International Conference on*. IEEE, 2014, pp. 1–8.

[45] L. Liu, N. Sang, S. Yang, and R. Huang, "Real-time skin color detection under rapidly changing illumination conditions," *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 3, pp. 1295–1302, 2011.

[46] R. Gopalan, S. Taheri, P. Turaga, and R. Chellappa, "A blur-robust descriptor with applications to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 6, pp. 1220–1226, 2012.

[47] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," in *Computer Vision–ECCV 2014*. Springer, 2014, pp. 47–62.

[48] Y. Dupuis, X. Savatier, J.-Y. Ertaud, and P. Vasseur, "Robust radial face detection for omnidirectional vision," *Image Processing, IEEE Transactions on*, vol. 22, no. 5, pp. 1808–1821, 2013.

[49] S. Gunasekar, J. Ghosh, and A. C. Bovik, "Face detection on distorted images using perceptual quality-aware features," vol. 9014, 2014, pp. 90 141E–90 141E–13. [Online]. Available: http://dx.doi.org/10.1117/12.2037343

[50] P. Korshunov and W. T. Ooi, "Video quality for face detection, recognition, and tracking," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 7, no. 3, p. 14, 2011.

[51] V. Jain and E. Learned-Miller, "Fddb: A benchmark for face detection in unconstrained settings," *Technical Report UM-CS-2010-009, Dept. of Computer Science, University of Massachusetts, Amherst*, 2010.

[52] C. Zhang and Z. Zhang, "Winner-take-all multiple category boosting for multiview face detection," *Microsoft Research TechReport, TR-2009-190, pp. 1-9*, 2009.

[53] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.

[54] C. Zhang and Z. Zhang, "Improving multiview face detection with multi-task deep convolutional neural networks," in *IEEE Winter Conference on Applications of Computer Vision*, 2014, pp. 1036-1041.

[55] Y. Zhiguo, G. Hao, P. Chun, and M. Lin, "The study on face detection strategy based on deep learning mechanism," in *Future Information Technology, Springer*, 2014, pp. 538-541.

[56] H. Li, Z. Lin, X. Shen, J. Brandt, and G. Hua, "A convolutional neural network cascade for face detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5325-5334.

[57] S. S. Farfade, M. Saberian, and L.-J. Li, "Multi-view face detection using deep convolutional neural networks," in *arXiv:1502.02766*, 2015, pp. 1-8.


Image Processing