



Image Enhancement for Impaired Vision: The Challenge of Evaluation

Eli Peli

The Schepens Eye Research Institute, Department of Ophthalmology, Harvard Medical School,

Boston, MA, USA

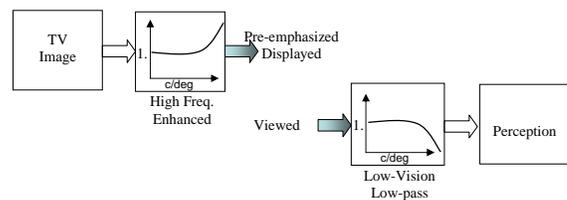
eli.peli@schepens.harvard.edu

With the aging of the population, the prevalence of eye diseases and thus of vision impairment is increasing. The TV watching habits of people with vision impairments are comparable to normally sighted people [1], however their vision loss prevents them from fully benefiting from this medium. For over 20 years we have been developing video image-enhancement techniques designed to assist people with visual impairments, particularly those due to central retinal vision loss. A major difficulty in this endeavor is the lack of evaluation techniques to assess and compare the effectiveness of various enhancement methods. This paper reviews our approaches to image enhancement and the results we have obtained, with special emphasis on the difficulties encountered in the evaluation of the benefits of enhancement and the solutions we have developed to date.

The pre-emphasis model

In our linear pre-emphasis model [2, 3], the loss of contrast sensitivity due to vision loss (which usually affect higher spatial frequencies more than lower frequencies [4]) is modeled as low pass filtering of the displayed image (Fig. 1). To counteract this effect, the displayed image is pre-emphasized by enhancing the higher frequencies before displaying the image [2]. However, due to constraints imposed by the display's dynamic range, only a moderate level of enhancement is possible. Further enhancement results in grey scale saturation [5, 6] and distortions. In addition, enhancement of high spatial frequencies beyond vision limits will not have any effect, and enhancement of low spatial frequencies that viewers can readily see does not bring any perception benefit. Therefore, enhancement of a limited band of frequencies that viewers otherwise would not be able to see is preferable. A number of enhancement techniques we have developed are summarized here, followed by reports of our evaluation studies. The challenge of evaluating image enhancement is the central theme of this paper.

Fig 1. The pre-emphasis model. The pre-emphasis filter designed to neutralize the effect of the low vision filter is applied to the displayed image prior to presentation. Dynamic range limits of the display restrict the direct application of the model.

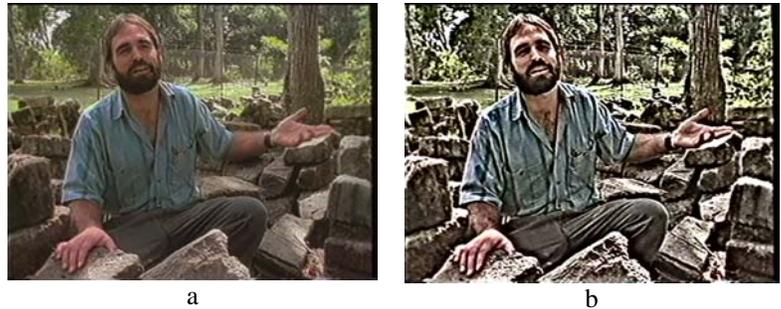


Enhancement Methods

1. Adaptive enhancement

The adaptive enhancement algorithm [7] that can enhance a tuned range of frequencies and can reduce saturation by reducing low frequencies was implemented, and tested first in software for static monochrome images [2, 4] and then in hardware for motion color video [8, 9]. Results of enhancement are illustrated in Fig. 2 for a frame taken from a video processed in real time using the DigiVision CE-3000 device and in Fig. 3b as computed for static monochrome images.

Fig. 2. (a) The original, unenhanced video frame and (b) A video frame enhanced using the DigiVision CE-3000 device. Note the moderate level of enhancement and changes in local luminance (particularly on the trouser leg) that permit greater enhancement of high frequency details (such as the folds in the material)

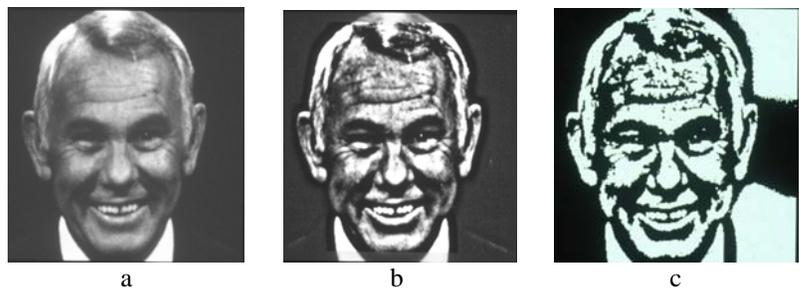


2. Adaptive thresholding

Adaptive or local thresholding that results in binary images is not commonly considered as an enhancement technique, but may serve as such especially for the visually impaired. The binary image has inherent high contrast and if it maintains the relevant image's information in a satisfactory way, it may be useful as an enhancement technique.

Adaptive thresholding changes the threshold applied across the image as a function of local image properties [10, 11]. The size of the local neighborhood used in processing the local properties determines the range of spatial frequencies that are enhanced by the adaptive thresholding. An example of a face image enhanced with adaptive thresholding is shown in Fig. 3c. While the face is clearly recognizable, the severe distortion caused by the enhancement is apparent and is noted by the patients.

Fig. 3. (a) A celebrity face image unenhanced. (b) The adaptive enhancement of the image in a. (c) The adaptive thresholding enhancement of the same image.



An early head mounted displays (HMDs) system, the Private Eye, was a binary display device that used scanning red LEDs to form the image [12]. We adapted it as a portable low vision image enhancement aid. To reduce the cost, weight, and power consumption, a one dimensional (1-D) analog video processing alternative was designed and implemented [13]. However, this approach was never formally tested to demonstrate its effectiveness since there was no clear way to assess the value of such system for mobility.

3. *Tuning the enhancement individually*

In both adaptive enhancement and adaptive thresholding, the band of spatial frequencies being enhanced can be selected and therefore may be adjusted for an individual user. The gain of the adaptive enhancement may be tuned as well. The pre-emphasis model suggests that better results might be achieved with individual tuning of the band. The significant and substantial increased face recognition for about half the patients we achieved with uniformly applied enhancement [4] suggests that tuning of the enhancement may result in even better results. The value of tuning is even more difficult to determine than the benefit of the enhancement. As a result of these difficulties and various other considerations relating to the effect of bandwidth of stimuli on visual function (especially in the peripheral visual field), we have developed and evaluated the use of wideband enhancement [14].

4. *Wideband Enhancement*

The wideband image enhancement consists of locating visually relevant features in the image (edges and bars) and enhancing the contrast of the pixels of such features [14]. The edge detection algorithm used for the wideband enhancement is a dual polarity edge detector based on a vision model [15] (Fig. 4). This algorithm marks “edge” features with dual polarity pairs of bright and dark lines with the bright line on the bright side of the edge and the dark line on the dark side of the edge. Thin “bar” features are represented with a single, appropriate polarity line at the location of the bar. The feature’s outlines detected by the algorithm may be used to enhance the visibility of the features they underlie. Bright and dark lines can replace (substitute) the original pixels values at their corresponding locations or they can be added (subtracted for dark lines) to the original pixels’ values. In both cases the outline magnitudes can be fixed or variable. Following a set of pilot experiments we selected a process whereby outlines detected in the image were added to the original image at their locations but were scaled in magnitude according to the strength of the feature at the location.

We have also proposed the use of the wideband enhancement in a see-through HMD as a way to provide augmented vision – by enhancing the view of the real world. In this application only unipolar (white) edges can be used as it is not possible to implement a black edge on the optical see-through display. It is critical in this application to achieve accurate registration between the natural view seen through the display and the edge images presented on the display. We, therefore, have developed and evaluated (in conjunction with MicroOptical Inc) a system in which the same optical path is used for the camera acquiring the images and for the display presenting the edges derived from it. While we were able to achieve the required registration [16], the brightness of the edges that could be achieved with the existing LCD display technology was insufficient to provide a beneficial enhancement effect. Other emerging technologies may provide better alternatives.



Fig. 4. Wideband enhancement. (a) Bipolar features (edges and bars) detected from the original image are scaled according to the strength of the feature. (b) The bipolar features added to the original image. Note the white line on the woman's left arm. The features brightness can be adjusted to increase or decrease the level of enhancement.

5. Enhancement of Compressed Video

The approaches described above were based on the filtering of analog (uncompressed) video. However, the use of digital video products applying MPEG compression, such as digital televisions, DVD players, and digital camcorders, is rapidly expanding. Techniques for video enhancement to aid visually impaired people must therefore evolve to be compatible with the new digital media formats. Enhancing MPEG compressed images often results in significant block artifacts which become visible for both normally sighted and visually impaired audiences; and pre-compression enhancement is impractical and defeats the compression efforts. We consequently developed an MPEG-based video enhancement that operates during the decompression stage [17-19]. This approach reduces the appearance of block artifacts and efficiently uses the decompression infrastructure. Our MPEG-based enhancement is based on the discrete cosine transform (DCT) and the quantization matrix of the JPEG aspect of the MPEG standard. The properties of the DCT provide a very natural way for defining spatial frequency filters in the frequency domain. Within the 8×8 block commonly used in MPEG and JPEG coding; the top-left function represents the "DC" or zero spatial frequency, along the top row the basis functions increase in horizontal spatial frequency, down the left column the functions increase in vertical spatial frequency, and along the diagonals there is an increase in both horizontal and vertical frequencies.

As set by the pre-emphasis model, effective image enhancement requires increasing the contrast in a specific range of frequencies. Applying filtering in the DCT domain can be achieved within the MPEG decompression stage by manipulating the Q matrices available in the sequence header (there are two different Q matrices — Intra and Inter matrices — with different values for quantization of still and moving macroblocks). In our enhancement approach, both the Intra and Inter Q matrices are multiplied, point-by-point, with pre-designed Intra and Inter enhancement filter arrays to obtain the modified Q matrices. This technique only requires access to the Inter and Intra quantization matrices being decoded from the header, and the ability to multiply them with the enhancement filter arrays.

We implemented this approach first with static images using custom programming of the JPEG decompression stage [18]. In that first study we applied a uniform enhancement factor at all frequencies (wideband enhancement). We later used the same approach for band-limited filtering applied to MPEG video test sequences with the MPEG software decoder [20]. The test sequences were processed off line and numerous versions with varying parameter settings were pre-prepared and stored on disk for testing [19]. Most recently the post transmission enhancement was integrated into an “open source” MPEG player that could process the video and adjust the parameter for live video fed from any MPEG source [21]. The flexibility of this system enabled improvement of the enhancement algorithm and development of reduction of interlacing artifacts (Fig 5).



Fig. 5. Illustration of ringing artifacts and their reduction in the MPEG enhancement [21]. (a) Unenhanced section of a movie frame with text. (b) Text is particularly prone to ringing artifacts caused by enhancement that is limited to a band of frequencies in the DCT block. (c) Our recent filtering techniques have substantially reduced these ringing artifacts without reducing the contrast enhancement effect. (d) Curved edges appear ‘jagged’ when ringing artifacts are caused by the DCT filtering. (e) The more recent filtering applied has reduced these artifacts.

Evaluation Methods and Results

1. Evaluation using simulations of vision loss

The potential benefit of the adaptive enhancement for visually impaired was evaluated first using photographic simulation of the effect of cataracts [2]. The original images and the enhanced images were photographed with a camera that was rendered cataractous by dubbing Vaseline on the camera lens with the finger. This treatment had been shown to be a good simulation of the optical effect of cataract in the eye [22]. This simulation of cataracts actually imposed a linear optical filter and thus it is not surprising that the pre-emphasis approach was effective and the simulations were judged promising [2].

Later we evaluated the Adaptive Enhancement and the Adaptive Thresholding using computational simulations of the loss of vision [4]. Two types of simulations were applied; a linear filtering (cataract), and a non linear processing approach that directly implements the threshold non-linearity of the human contrast detection system [23]. The latter was aimed at simulating the loss of contrast sensitivity in the retinal periphery which might represent the visual function of patients with central field loss (CFL) due to age-related macular degeneration (AMD) and other diseases that damage the fovea. The validity of these

simulations for both central foveal vision [24, 25] and peripheral vision [4] was confirmed experimentally. The simulations were again promising as judged in a side-by-side comparison of the results of simulating the degraded vision with and without the enhancement of the image. While such side-by-side visual comparison is a frequently practiced approach to assess the value of image processing techniques and image enhancement in particular, its value is clearly limited (see ref [26] for review). Simulation, nevertheless, could be a valuable tool in the process of developing enhancement algorithms and could be of benefit in initial testing and parameter settings. However, a direct testing of the effect of enhancement in people with impaired vision is essential in order to prove the benefit of the approach and of any specific technique.

2. Evaluation of performance

Intuitively it appears that if image enhancement is effective it should improve the ability of the person with visual impairment to perform certain visual tasks. The same approach is or should be true for some applications of image enhancement, e.g. medical images. Specifying relevant task performance to be tested is relatively straightforward for medical imaging. Usually it is a detection performance on some diagnostic imaging test. With a ground truth that may be established using biopsies or follow up, the ability to diagnose (i.e., correctly detect or identify lesion) using the original radiology image or the enhanced image can be compared. When the state of such evaluations was reviewed [26], there was little indication in the literature that image enhancement had improved diagnostic performance. This was perhaps due to the excellent ability of the normal visual system to use high quality displays to retrieve all necessary information from the original image. However, when a visually impaired system is used there is clearly a loss of information, which potentially may be recovered, at least in part, with the help of image enhancement.

The main difficulty in applying the same approach to assess the benefits of image enhancement to television viewing by people with vision impairment is the lack of clearly defined tasks to be evaluated. The difficult question is: how to define and qualify/quantify the task performed while viewing TV for pleasure and how would we measure the person's performance?

a. Evaluation of performance with static images: Face recognition

One of the most frequent complaints of visually impaired patients is the reduced ability to recognize faces both on TV and in real life. In daily life such failures may lead to embarrassing social interactions, and in TV viewing failing to recognize a person in a scene may affect the patient's ability to follow the story line. We therefore first used face recognition tasks in investigating the effects of image enhancement [4, 6]. Sergent [27] reported that although low frequencies convey most of the relevant information for face processing, high frequency information is not redundant. High frequencies seem to benefit the performance of tasks that require accessing the identity of a face as compared with discriminating among a small sample of face images. We chose celebrities' face recognition, a task that has been shown to be more robust than the recognition of

unfamiliar faces [28]. But as we found, it is not trivial to determine who is a celebrity and that ground truth needs to be established for the population under test or for each subject. For one population we tested (Americans with normal vision older than age 60 in the late 1980s), Mick Jagger was not recognized by anyone while Johnny Carson was recognized by all. Forty patients with CFL in one eye (VA worse than 20/70) due to macular disease participated in the celebrities' face recognition study [4]. All had VA better than 20/40 in the good eye that was used to verify the familiarity with the celebrities. Twenty-one of these patients were tested with the adaptive enhancement technique and nineteen with the adaptive thresholding. Photographs of 50 celebrities and 40 unfamiliar people were presented. Monochrome grayscale images digitized at a resolution of 256x256 and at 256 gray levels spanned 4x4deg at the viewing distance. The images were enhanced with the Adaptive Enhancement algorithm and the Adaptive Thresholding technique using the same parameters applied for the simulations evaluation. Only one of the enhancement techniques was used for each subject. Original (unenhanced) and enhanced images (180 in total) were presented in random order. Subjects indicated their level of confidence, on a scale of 1 to 6, in recognizing the face as that of a celebrity. A rating of 1 meant that the subject was positive that the face belonged to a celebrity; 6 meant that the face was clear but not recognized. Celebrities that were not recognized in both enhanced and unenhanced modes were presented to the patient's better eye. If the patient could not recognize a particular celebrity with his good eye that celebrity was reclassified as an unfamiliar person for the patient.

Celebrity detector ROC

Receiver operating curves (ROCs) plotting the probability of true celebrity vs. the probability of false celebrity were constructed from the responses, treating the patients as celebrity detectors. Separate curves were calculated for original and enhanced images. Because the same faces were presented in both forms, a correlated ROC analysis [29] was conducted. The area under the ROC (A_z) was taken as a measure of recognition [30]. Note that this is a standard application of the ROC technique which requires the establishment of the ground-truth (true celebrity in this case). The ROC curves for one patient are shown in Fig 6.

Most of the patients (33 out of 40) demonstrated improved face recognition with the enhanced images as compared with the original, unenhanced images. The difference between the two areas under the ROC curves indicated a statistically significant ($p < 0.05$) increase in recognition for 11 out of the 21 patients tested using the Adaptive Enhancement algorithm (empty triangles in Fig. 7). For 2 patients in this group recognition decreased with the enhancement but the difference was not significant. Six out of the 19 patients tested with the adaptive thresholding technique showed significantly increased recognition for the enhanced images and one had a significantly decreased recognition. To summarize the data for all subjects we defined an

improvement measure as the ratio of areas under the curves ($A_z(enhanced)/A_z(original)$). Results of the 21 patients in Adaptive Enhancement group as well as the maximal possible improvement shown in Fig 7.

Fig. 6. ROC curves for a patient who performs much better as a celebrity detector with the Adaptive Enhancement. The data points were derived directly from the patient's response. First data point on the left was calculated from responses of confidence level 1, and the other data points were cumulative probabilities. The lines represent the ROC fittings to these data points. In this case the difference between the two curves was statistically significant.

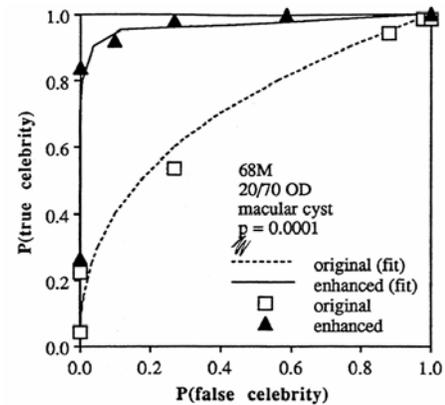
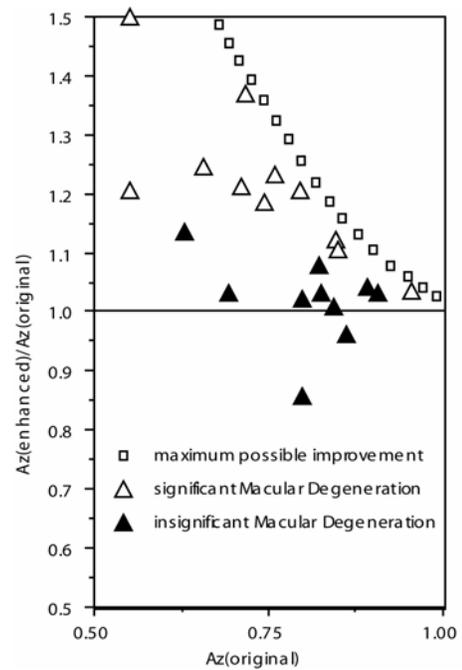


Fig. 7. Improvement in face recognition of patients with CFL with the Adaptive Enhancement as a function of performance without enhancement. Open triangles represent subjects with CFL that improved their performance significantly with the enhancement. Filled triangles represent subjects that had a change in performance with the enhancement but the change was not statistically significant. Only 2 subjects had a decreased performance with enhancement (triangles below the midline). Open squares represent the maximum possible improvement available for each level of performance with no enhancement. For some subjects the improvement approached the maximal possible improvement.



A normalized measure of improved performance, *gain*, was calculated as the ratio of increase in area under the ROC and maximal possible increase:

$$gain = \frac{A_z(enhanced) - A_z(original)}{1 - A_z(original)} \quad (1)$$

The mean gain was 62% of the maximal possible improvement for the patients that demonstrated a significant improvement with the Adaptive Enhanced images and 42% for the patients that showed significant improvement with the Adaptive Thresholding. These results were encouraging and demonstrated a meaningful benefit for most of the patients.

b. Evaluation of performance with motion videos

We know of no accepted or even suggested method to measure performance with video sequences. The difficulty is inherent to constantly changing content and image quality in a video sequence. This continuous change makes it difficult to capture a fleeting task performance. While in some situations a specific task could be defined (e.g., the detection of suspicious object in passengers' luggage in airport security scanning system), defining a performance measure relevant to TV viewing for pleasure or for information is much more difficult.

Peli et al [31] developed a performance measure to assess the effectiveness of Audio Description (AD) for the blind and visually impaired. AD provides verbal descriptions of the visual contents of TV programs through the third audio channel without interfering with the programs' standard audio portion [32]. Descriptions of visual details concerning aspects such as clothing and colors are inserted during pauses in the dialogue. AD is available on some DVDs and videocassette tapes, on some public broadcast programs in the USA, and in the United Kingdom. Peli et al [31] constructed questionnaires probing details described in the AD of three public broadcast programs. The effect of AD was evaluated by addressing these questionnaires to visually impaired subjects who watched the programs with or without the AD.

We applied the same questionnaires to testing recognition from the video program. We counted the number of visual details that could be correctly identified by subjects with impaired vision in response to the questions after observing either the original or the enhanced video segments (both played without AD) [9]. Multiple-choice questions were posed after each short segment. Questions addressed visual details, e.g. "The woman has... a) gray hair; b) black hair", that were described by the original AD prepared for this TV program for broadcasting on Public Television. The video program "Poirot: The Theft of the Royal Ruby", an episode of the Public Broadcast Service show "*Mystery!*", was used.

The questionnaire consisted of 59 questions covering a 10 minute segment of the episode. The video was paused 17 times, at proper break points, to administer the AD-based questions. The initial condition (either enhanced or unenhanced) was counterbalanced across subjects. After the 30th question, the condition was switched. Thus, half the subjects viewed the first part of the segment in the enhanced mode and half the subjects viewed it in the unenhanced mode (Fig. 2). The parameters for the enhanced condition were the individually-selected enhancement settings (see below).

The subjects (n=25) answered 66% of the questions correctly when the video was enhanced and 71% when it was presented with no enhancement. This difference was not statistically significant (paired sample t-test, $t_{24}=2.04$, $p=0.053$). Note that in agreement with our previous results [31], the subjects could answer over 70% of the questions correctly without enhancement and without hearing the DVS description, leaving very little room for potential improvement and also indicating that the AD was not appropriate for visually impaired people as they could determine 70% of its content without it. It is possible that a more adequate questionnaire (not based on the AD) could be developed to measure TV-watching performance but it is not

straightforward how to construct such questionnaire. Even if a questionnaire was constructed that could test recognition of visual details in the TV program that are missed or not visible to patients with impaired vision, it is not clear that such details are important or even relevant to the enjoyment or benefit of the viewer.

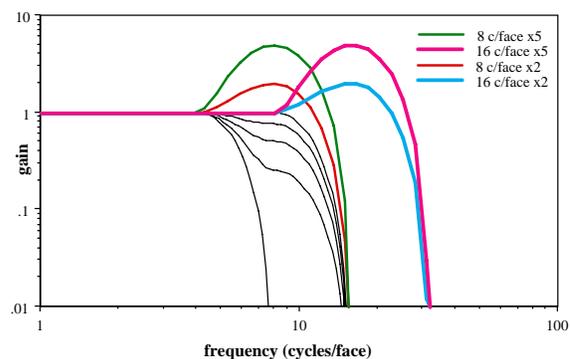
3. Individual Tuning as an Evaluation

If individual patients could tune the enhancement level consistently (in terms of spatial frequency, gain, threshold level or any other parameter) they would be demonstrating a perceived value of the enhancement (a dose effect). A number of studies were aimed at examining the effect of tuning on the enhancement for an individual patient's visual loss, but a clear benefit from individual tuning has not been found.

A pilot study using the DigiVision CE2000 device found increased recognition of details in the videos and almost all (95% of the subjects) preferred their individually-selected enhancement [4]. A different study, using a face recognition task, found that individually-selected enhancement parameters did not improve recognition more than uniform enhancement [6].

A study of live video enhanced with the DigiVision device, using fixed enhancement parameters and individually-selected viewing distance [33], found a statistically significant but relatively small improvement in performance, and only 20% of the subjects in the study indicated a preference for the enhanced images. In another study we explicitly applied a set of filters in the frequency domain to face images (Fig. 8) and thus directly controlled the frequencies that were enhanced as well as the bandwidth of the range of frequencies being enhanced [6]. The wider bandwidth was achieved by combining two narrow filters. We showed that the effect of a 2-octave wide filter with a gain of $5\times$ was very similar to the effect of the adaptive enhancement algorithm applied in the previous study [4]. The low pass filtering was applied to determine the critical frequencies needed for face recognition (4-8 cycles/face). Patients selected their preferred enhancements from two pallets of 4×4 pre-computed images selected by moving a mouse over a graphic tablet (total of 32 possible selections). The selected enhancement and a uniformly applied enhancement were then applied to the face images used in the celebrity detection task. We found that patients preferred enhancement at frequencies higher than the critical frequencies. They also preferred the wideband enhancement over the narrowband enhancement. Individually selected enhancement, however, did not improve face recognition in comparison to uniformly applied enhancement [6]. The same approach of asking the patient to select from a pre-computed set of images the preferred one using a mouse to select the various options was applied also in tuning the gain of the wideband enhancement [14]. Only 5 out of 35 patients selected the original unenhanced image and none selected the degraded images that were included in the selection, indicating that the enhancement was preferred. Most patients selected a moderate level of enhancement and only few selected high level of enhancement. Similar tuning of the level of enhancement was also used to select the individual's level of enhancement in the study of the enhancement in the JPEG domain [18].

Fig. 8. The set of lowpass (thin lines) and band-enhanced filters (color thicker lines) used to evaluate the effect of tuning on the enhancement preference and performance [6]. Note that that while the lowpass filters are conventional, the enhancement filters are not simply band-pass but specifically band-enhanced filters with gain>1. The notation x5 refers to a gain of 5.0 for that filter



In a more recent study we evaluated the effect of the adaptive enhancement applied to motion video using the DigiVision CE3000 device, again assessing the effect of individual tuning of the enhancement [9]. In pilot experiments we had found that patients (and also investigators wearing a cataract simulating glasses) had great difficulties setting the enhancement parameters for motion videos. Because of the changing nature of the video content there was no way to determine a single setting that was optimal or satisfactory for the whole video. The observers felt that the setting had to be flexible. We therefore implemented the parameter selection step again using static frames captured from the videos (Fig. 9a). The subjects used mouse to control the ‘Detail’ parameter, representing spatial frequency, along the up/down dimension and a combination of the ‘Contrast (Background)’ parameters in the right/left dimension.

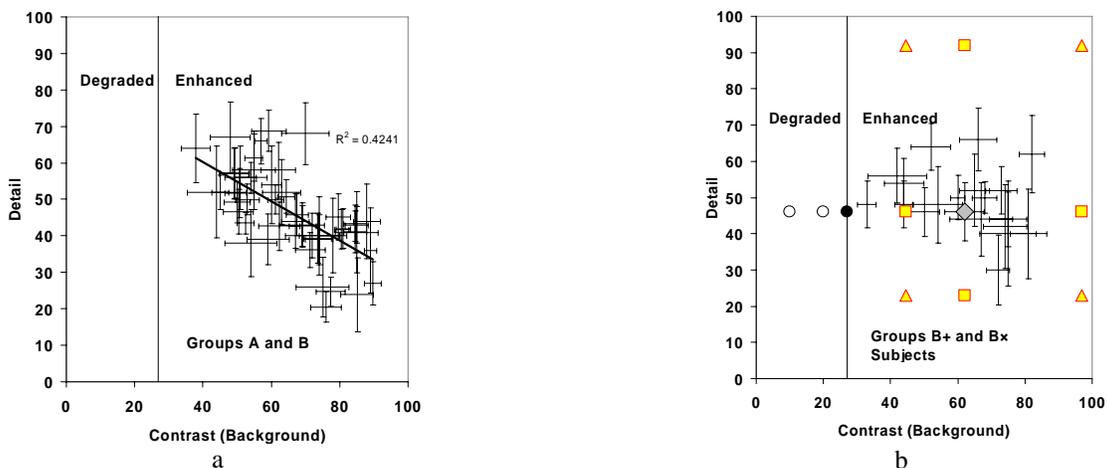


Fig. 9. Enhancement parameters chosen by each subject. The position of each cross corresponds to mean mouse position on the graphics tablet and the error bars represent SEM. The vertical line at contrast = 27 (arbitrary units) represents no enhancement (original image). Area to the left of that line represents image degradation by low-pass filtering. Area to the right represents enhancement. All subjects selected values corresponding to enhancement. (a) Mean settings and standard deviations for each of the 49 patients who viewed static images, (b) Mean settings of 20 subjects who completed the continuous evaluation of motion videos. The filled symbols to the right of the vertical line represent enhancement settings used in the second part of the experiment for one subject. Diamond: individually selected enhancement settings; Triangles: settings used if in the B \times group; Squares: settings used if in the B+ group; Open circles: settings that resulted in two degraded images; Filled circle: setting for the unenhanced original image that is produced from any detail setting that had contrast =27. For static images, visually impaired subjects could clearly select a preferred level of enhancement repeatedly, as indicated by the small error bars. As a group, they showed a clear pattern of selecting a higher contrast gain setting if they selected a lower spatial frequency band for enhancement (Fig. 9a). However, when we continuously tracked the perceived quality of motion

video, using both the individually-selected and modified enhancement parameters (Fig. 9b), we found that subjects significantly preferred all of the enhancement parameter choices. The subjects' individually selected enhancement parameters resulted in the largest effect, although this was not significantly better than the other enhancement options [9]. Thus, the value of enhancement tuning still remains elusive.

a. Individual Tuning of Movie Sequences

While selecting a single setting for the whole segment of video is difficult, continuous adjustment of setting is possible. It is easier to perform this task when only one parameter is varied. We have recently used this approach in evaluating the benefit of the MPEG enhancement performed in real time. Subjects were presented with short (4 minutes) video segments of 4 different styles and were asked to continuously adjust the enhancement parameter using the up and down buttons on a TV remote control. Every 2 minutes the enhancement parameter was set to an extreme value: either a high level of enhancement that resulted in clearly noticeable distortions or a very high level of compression resulting in very blurred images. Following each such change the subjects were asked to use the up and down buttons to search for the best (clearest) view. The size of the change with each button press was changed from 4 just noticeable difference (JND) steps at first to 1 JND after the second reversal in staircase direction. If the subject had not adjusted the display for 15 sec, a voice prompt requested readjustment of the setting. Most patients appreciated the enhancement and were able to set the level consistently to a preferred level (Fig. 10a). A few patients with severe vision loss could not appreciate any effect of this enhancement (Fig. 10b). The data collection for this study (24 subjects) has just been completed and a complete analysis is not ready yet. The results however are very encouraging as they appear from examining the raw data.

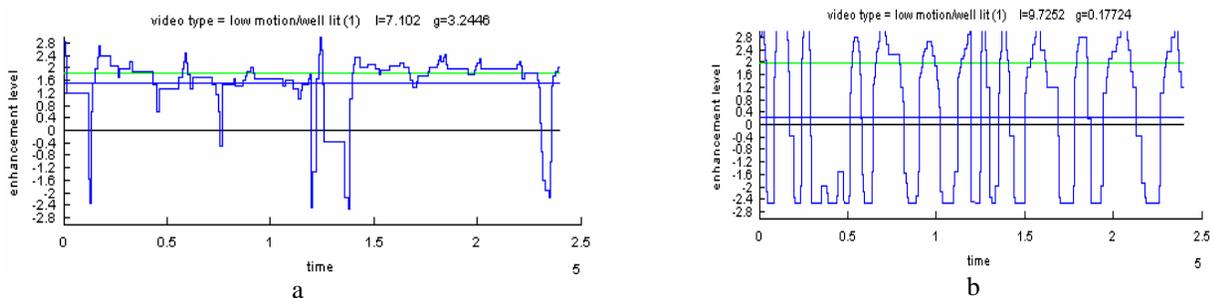


Fig. 10. MPEG enhancement settings while watching 4 minutes of a low motion well lit video (a panel discussion). The zero level (black horizontal line) represents the original image and negative values represent degradation. The other horizontal lines represent linear (green) and geometric (blue) means of the settings. (a) The remote control setting of a patient with acuity of 20/250 who clearly appreciated the effect of enhancement. The segment starts with over enhancement and is reset at the middle to very low setting. In both cases the patient is resetting the enhancement to a level of about 1.6 to 1.8. (b) The setting selected for the same video segment by another patient with more severe loss (20/600) who did not appreciate the effect of the enhancement (or degradation) at all. Settings are arbitrarily moving between the extreme values. Time axes in samples $(30/\text{sec}) \times 10^5$.

4. Evaluation of impression of quality

The simplest approach to determining the impact on image quality is to ask the viewer for their impression of the quality. The subject may be asked to rate the quality on a verbal scale such as “poor, average, good, and excellent” or to assign a numerical value to it. This approach can be implemented to single frame images when evaluating enhancement of static images or to the evaluation of a video sequence following the presentation of a processed sequence.

a. Evaluation of impression with a video sequence

We have applied the latter approach in one of our studies [9], where each subject was asked to mark his response to 7 questions, comparing video segments just seen to normal TV viewing. The responses were indicated by moving a marker across a continuously numbered scale, which was labeled by the words “poor” and “excellent” at the ends of the scale in large print. The experimenter recorded the subject's responses from the scale (range of 0 to 50). Comparisons were made of the seven measures: color, visibility of details, ability to recognize faces, ability to discern facial expressions, ability to follow the story, sound quality, and overall impression. Following the presentations of the second of two conditions in the performance study, the comparison questions were repeated. In this case, for each question the experimenter positioned the marker to the previous setting selected by the subject for this question and the subject was asked to indicate their response in comparison to the previous selection for the first condition. The overall mean of these scores was 0.15 ± 0.08 (SEM), which indicated a slight preference for the enhanced images but it only approached statistical significance (one-sample t-test, $t_{111}=1.89$, $p=0.062$). Subjects remarked that it was difficult to make the required comparison to “normal TV” and to the other segment as the two segments differed in content even though they were continuations of the same program.

b. Evaluation of impression with static images

A more elaborate evaluation of the impression of image quality using static images was applied in analyzing the effects of two types of enhancement: the JPEG enhancement [18] and the wideband enhancement [14]. Both studies used the same 50 static TV images sampled from a TV cable broadcast. In each study another set of images from the same source was used to select the individual preferred level of enhancement. Once the individual level was selected for a subject, a set of 200 images was created consisting of 4 versions of each image to be used in the quality evaluation study. The four versions included: (1) original image; (2) individually chosen enhancement; (3) a degraded image; and (4) an image enhanced by a second arbitrarily selected enhancement level. The images were presented to each subject in a random sequence. The patients were asked to rate the image as “better”, “slightly better”, “typical”, “slightly worse” or “worse” than the original image by moving the mouse on the graphics tablet. These words were printed in a large font on the graphics tablet. Before the computer accepted their rating, the patients were forced to view the original image at least once for comparison.

Image quality detector ROC

The data from these two studies using static images and a third study that employed the continuous evaluation of impression of image quality were analyzed using a somewhat different ROC approach. Paired comparisons were made between responses to the original images and processed images. As there were three sets of processed images for each patient, three ROC curves were determined (Fig. 11). These represented the difference in perceived image quality between the original and processed images.

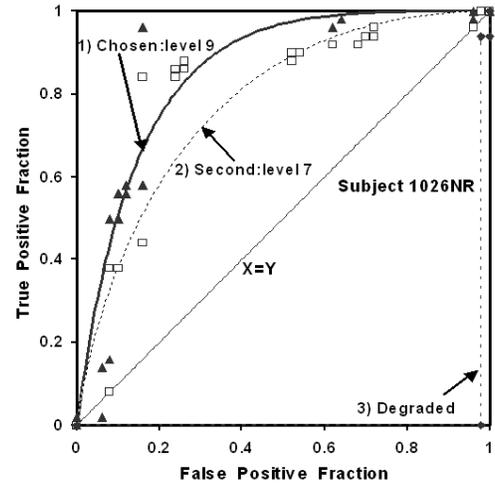
In standard ROC analysis (as described above for the adaptive enhancement study), a detector's (e.g. patient's) responses to "noise" (non celebrity) presentations and to "noise-plus-signal" (celebrity) presentations are compared. In such standard application of ROC there is a ground truth to which the responses are compared. In our impression of image quality studies, the original images were treated as the noise presentations, and the processed images were treated as the noise-plus-signal presentations. Patients were asked to report perceived image quality, so they could be considered image-quality detectors, comparing the quality of images presented with one set of parameters to the quality of the original unenhanced images. The raw data consisted of multiple frequency distributions along the perceived image quality dimension. When the perceived image quality of the processed images was better than the original images, A_z was greater than 0.5. For the degraded image set, subjects' perceived image quality distributions were always worse than that of the original images, resulting in $A_z < 0.5$. As our ROC analysis was of perceived image quality - not of enhancement detection, as might be done in another application - the traditional labels of the axes of the ROC figure (e.g. true-positive rate, or "hit" rate) do not apply directly to our situation. In our analysis, the true-positive rate dimension was the proportion of the *processed* image set with a higher perceived image quality, while the false-positive rate ("false-alarm" rate) dimension was the proportion of the *original* image set with a higher perceived image quality (higher being relative to the criterion used for that point on the ROC curve).

The results for the JPEG enhancement were not indicative of a significant impression of improved quality [18]. Only a third of the patients' data indicated any preference for the enhancement and for most the difference was not significant. The JPEG based enhancement algorithm has been improved twice since then and testing with the improved algorithms has shown much better results than what was reported above in (Individual Tuning of MPEG Movie Sequences) and below (Side by side evaluation of MPEG enhancement) [19].

Twenty-three patients participated in the ROC evaluation of image quality for the wideband enhancement. For the 19 patients who preferred the wide band enhancement, the individually-selected wide-band enhancement was found on average as having slightly better image quality than the original images ($A_z = 0.57 \pm 0.026$; $p = 0.012$). Five of the 23 patients (22 %) had, an A_z significantly greater than 0.5 (i.e., $A_z \geq 0.68$) and 3 other patients approached this level of significance. Thus for the majority of the patients the improvement provided by the wide band enhancement was not significant. Possible reasons for these results

are discussed in ref. [14]. Here I would like to highlight that the evaluation methodology we used could determine and quantify even the very modest effect of the wideband enhancement as applied in our study.

Fig. 11. ROC data and fitted curves for a patient in the wideband enhancement study. True positive is the proportion of the processed images with higher perceived quality and False positive is the proportion of the original images with higher perceived quality. The thick solid line is the fit to the filled triangles (individually-selected enhancement level) and the thin line is the fit to the open squares (second arbitrary level). The dashed line hugging the lower right corner is the fit to the filled diamonds (the degraded images). This 43-year-old patient (VA 20/250) clearly rejected the degraded image and significantly favored the enhanced images.



c. Continuous evaluation of impression of quality

The difficulty observers have with evaluating the image quality of a video is affected both by difficulties in comparing image quality across different images in time and on the need to integrate an impression that may be varied substantially due to the image quality of the unenhanced video which may change from scene to scene. In order to address both difficulties we have developed a method of continuous evaluation of quality derived from the method that Hamberg and de Ridder [34] used to evaluate perception of dynamic changes applied to static imagery.

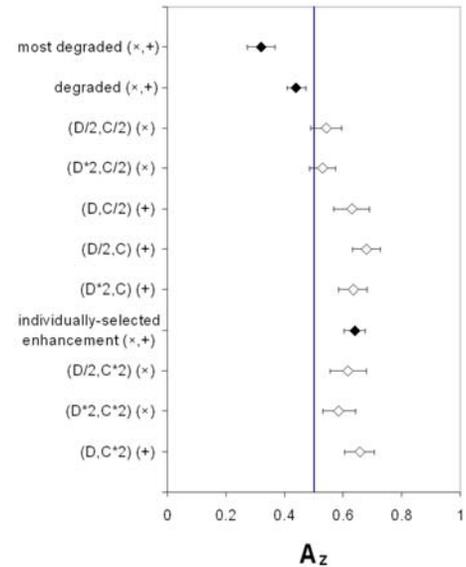
In this method, subjects indicated impression of image quality while viewing a movie: (**excellent, good, sufficient, poor, and bad**), by moving the mouse on a scale printed in large print. An auditory cue (beep) every 10 seconds indicated to the subject a change in parameters of the DigiVision adaptive enhancement. Mouse position selected in response to the new parameters was recorded (once per second). Data from the last 7 seconds in each 10-second interval were collected and averaged.

Two groups with 10 subjects each (Groups B+ and B×) participated in the study. For both groups each subject was presented with: her/his individually selected enhancement, the original unenhanced segments, and two different levels of degraded images. Each subject was also presented with sets processed with 4 additional arbitrary enhancement levels, two of which were over-enhanced. For one of these groups the plus (+) configuration of the arbitrary enhancement parameters was used, and for the other group the crossed (×) configuration was used (Fig. 9). Each condition was repeated 10 times and the scores for each of these repetitions were converted to the probabilities used for the ROC analysis. This was the same image quality ROC analysis as described above.

The results for the individually-selected enhancement and the degraded images were averaged for the two groups, while the arbitrary conditions were averaged separately for each group. The results (Fig. 12) demonstrate that subjects preferred the

enhanced videos to the unenhanced videos ($t_{107} = 6.92$, $p < 0.0005$) and preferred the unenhanced videos to the low-pass filtered videos ($t_{41} = -4.06$, $p < 0.0005$). The average rating for the original is equivalent to the “Sufficient” setting and the average rating for the individually selected enhancement is equivalent to the “Good” setting. Additionally, individually selected enhancement resulted in statistically significant improvement in perceived quality ($A_z = 0.64$, ± 0.17) over the unenhanced images (0.5) (one sample t-test, $DF=21$, $p=0.001$). No differences in perceived quality were found between the individually selected set of parameters and the corresponding arbitrary enhancement values in either group.

Fig. 12. Average A_z of 20 subjects. The three filled points show the two degraded conditions and the individually-selected enhancement condition, which were common across all 20 subjects. The degraded conditions have average A_z less than 0.5 whereas the enhanced condition has an average A_z greater than 0.5. The other points show all the other enhanced conditions for the (B_+) and the (B_\times) groups of 10 patients each. For all except two of these enhancement conditions, the A_z is significantly greater than 0.5 as indicated by lower bound of the error bars (SEM).



5. Evaluation of preference

Perhaps the most direct way to measure the value of video enhancement is to ask the observer (normally sighted or visually impaired) to indicate a preference in a direct side by side comparison of the enhanced and unenhanced images. If numerous comparisons of this type are repeated over multiple video sequences employing a proper study design that counterbalances the side of the presentation and with multiple observers, it should be possible to quantify preference. We also argue that for the purpose of enhancing TV for personal enjoyment and improved viewing experience such preference measurement may be more relevant to the value of enhancement than any other task performance improvement. We have conducted two studies of video enhancement using this evaluation approach; one for the MPEG method [19] and one for the adaptive enhancement [35].

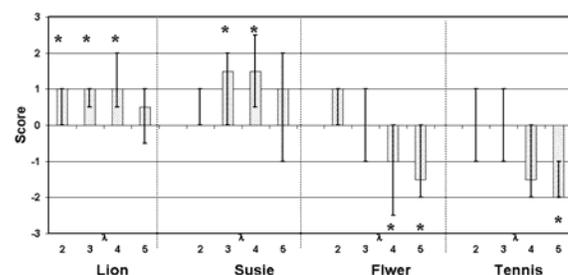
a. Side by side evaluation of MPEG enhancement

In the study evaluating the MPEG enhancement using off-line processing of test sequences [19], we enhanced and cut each video sequence to half the original width, but maintained the center of the picture. We then merged the original and enhanced

sequences using mirror-reversed replacement of the half-video to enable the side-by-side comparison of similar image areas. Both left-enhanced and right-enhanced sequences were created to allow balancing of the side on which enhancement was presented. A total of 32 video sets (4 sequences x 4 gains x 2 sides), each 5-seconds in length, were generated this way. Subjects sat at approximately 36" from a 19" PC monitor (1600 x 1200) and were asked to evaluate each side of the video sequence for "how clear the video is, how much detail and information could be obtained from it, and how is the general quality of the picture?" Using these guidelines, they were asked to choose which side of the video (left or right) they preferred. Subjects were forced to choose a side and the 5-seconds video sequences were repeated until the subject responded. Once they chose a side, they were asked to rate the chosen side relative to the other side as "a little better," "better," or "much better" than the other side (responses were recorded as a score of 1, 2, or 3). If a subject selected the enhanced side sequence, a positive score was assigned. If the subject selected the original un-enhanced sequence, a negative score was assigned. The negative or positive score from the first question was combined with the second question to yield a score that ranged from -3 to 3 except zero. Two scores were derived from each level of enhancement for each sequence (one score when enhancement was on the left side and one when it was on the right). The two scores were averaged.

Twenty-four visually impaired subjects (14 men), median age 71.0 years, with visual acuity ranging from 20/70 to 20/2500 participated in this study. All subjects had documented CFL in both eyes. During the experiments we noted that a few subjects seemed to have a clear preference for one side of the screen. We therefore tested for each patient if the selection was the same for the two identical presentations that differed only in side. For 11 of the 24 subjects, the preference was dependent on the side of the display (Paired t-test, $P < 0.05$) indicating a bias to one side, and therefore they were excluded from analysis. Fig. 13 shows the results from the remaining 13 subjects who showed consistent (unbiased) preference. The results of these 13 subjects were similar to those of the whole group. The two sequences that did not benefit from the enhancement were interlaced video sequences in which the enhancement substantially increased the interlacing artifacts.

Figure 13. The median values of the 13 subjects who did not show a bias to one side of the screen. The subjects preferred the 3 lower enhancement levels ($\lambda = 2$ to 4) for two of the sequences ("Susie" and "Lion"). The two highest enhancement levels ($\lambda = 4$ and 5) for the two other sequences ("Flwer" and "Tennis") were rejected. The lower enhancement levels were not significantly different from the original.

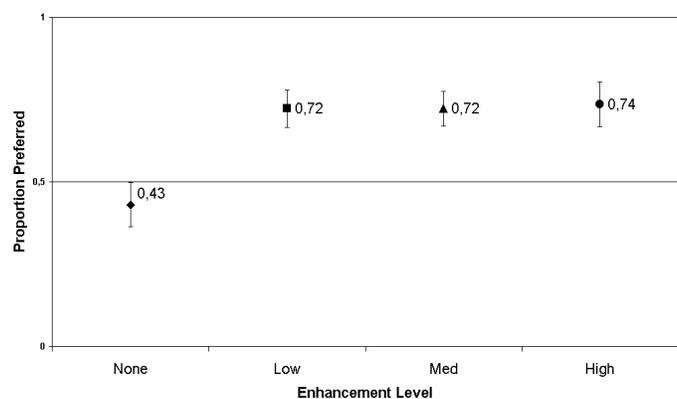


b. Side by side evaluation of adaptive enhancement

In a recent study [35] evaluating image enhancement with the adaptive enhancement as developed for a consumer home theatre product, we presented DVDs sequences from seven short video clips selected to represent various programs typically seen on

TV. Twenty subjects with CFL, seated 3 feet from and centered between two identical 27-inch televisions. One television picture was processed using the DigiVision device (representing the setting of the Belkin's RazorVision commercial product), the other was not. Subjects were asked 16 times which television movie looked 'clearer' to them. Each television received the processed video for half of these presentations, and this was repeated twice. The level of enhancement processing controlled by the counterbalancing table gave 4 presentations of each level ('bypass', 'low', 'medium' and 'high') per subject. A custom software program enabled the experimenter to record the subject's choice of clearer image, started video playback at the correct video, controlled the video switch for correct output, and sent the appropriate key presses to the DigiVision control software to select the correct enhancement level. Nineteen subjects, median age 60 years, visual acuity ranging from 20/46 to 20/609 (16 of them had documented CFL) took part in this study. Two subjects showed a strong preference for one side (75% in both cases, $p=0.028$). Data from these two subjects were removed from general analysis. The remaining data were pooled across all subjects and presentation side and analyzed using the proportion of presentations preferred. Fig. 14 shows a clear and significant preference for enhancement (Chi-Square=21.0, $p<0.001$).

Fig. 14. The overall proportion of processed presentations that were preferred by 17 subjects for each enhancement level, including the 'No Enhancement' bypass level. This bypass level is close to 50%, as expected. Error bars show \pm standard error of the mean.



Discussion

Image enhancement for improving video-viewing enjoyment of people with vision impairment is a promising approach and currently it is the only alternative to magnification approach (although it can work in conjunction with magnification). Current technological development makes an idea conceived over 20 years ago a potential reality in terms of capabilities and cost. This is particularly evident following our most recent results evaluating the DigiVision device, which implemented the parameter settings of a commercial product available for \$250. The transition to compressed digital video represents new challenges and new opportunities for the application, and if successful could become integrated in digital TV systems at essentially no additional cost for the user.

The only obstacle remaining for achieving this transition is a clear and convincing demonstration of the value derived from the enhancement by people with visual impairment. The difficulties encountered when trying to evaluate the preference and performance of enhancement are similar to those facing any other attempts to improve video technology, from HDTV to high

brightness displays. The essential difficulty is how to create a clearly quantifiable measure of image quality that can be obtained from an observer watching a video program on TV. The challenge is particularly difficult if one attempts to measure an improvement in the performance of any (preferably relevant) task in visually impaired. This has proven elusive as it is not clear what constitutes such a task since people watching TV programs are doing so for their entertainment. Measuring preference and/or impression of quality of a live motion video has also proven difficult as the image quality of the underlying images vary with time making a judgment difficult and variable. We have investigated numerous approaches and feel that two approaches seem most promising. First, the selection of parameters for a preferred setting in response to a shift of a parameter to an extreme value is a potentially workable method. We have not yet attempted addressing more than one parameter at a time under this approach and we do not yet have sufficient experience to determine the stability and repeatability of these measures. Second, side by side comparison of a live video, while its parameter settings are changing may serve as a reasonable approach. This approach has the advantage of simplicity and the potential for testing each subject over sufficiently large number of repetitions over a variable program material to let an average preferred setting or enhancement approach emerge. We plan to continue investigating these approaches while we search for even more effective enhancement algorithms.

Acknowledgements

Supported in part by NIH grants EY05957 and EY12890

References

- [1] J. Packer and C. Kirchner, "Who's Watching? A Profile of the Blind and Visually Impaired Audience for Television and Video," American Foundation for the Blind, 1997.
- [2] E. Peli and T. Peli, "Image enhancement for the visually impaired," *Optical Engineering*, vol. 23, pp. 47-51, 1984.
- [3] E. Peli, L. E. Arend, Jr., and G. T. Timberlake, "Computerized image enhancement for low vision: New technology, new possibilities," *Journal of Visual Impairment & Blindness*, vol. 80, pp. 849-854, 1986.
- [4] E. Peli, R. B. Goldstein, G. M. Young, C. L. Trempe, and S. M. Buzney, "Image enhancement for the visually impaired: Simulations and experimental results," *Investigative Ophthalmology & Visual Science*, vol. 32, pp. 2337-2350, 1991.
- [5] E. Peli, "Limitations of image enhancement for the visually impaired," *Optometry and Visual Science*, vol. 69, pp. 15-24, 1992.
- [6] E. Peli, E. Lee, C. L. Trempe, and S. Buzney, "Image enhancement for the visually impaired: the effects of enhancement on face recognition," *Journal of the Optical Society of America A*, vol. 11, pp. 1929-1939, 1994.
- [7] T. Peli and J. S. Lim, "Adaptive filtering for image enhancement," *Optical Engineering*, vol. 21, pp. 108-112, 1982.
- [8] R. G. Hier, G. W. Schmidt, R. S. Miller, and S. E. DeForest, "Real-time locally adaptive contrast enhancement: A practical key to overcoming display and human-visual-system limitations," presented at 1993 SID International Symposium Digest of Technical Papers, Seattle, WA, 1993.
- [9] E. Peli, "Recognition performance and perceived quality of video enhanced for the visually impaired," *Ophthalmic and Physiological Optics*, vol. 25, pp. 543-555, 2005.
- [10] C. K. Chow and T. Kaneko, "Automatic boundary detection of the left ventricle from cineangiograms," *Comput. Biomed. Res.*, vol. 5, pp. 388-410, 1972.
- [11] E. Peli and M. Lahav, "Drusen measurements from fundus photographs using computerized image analysis," *Ophthalmology*, vol. 93, pp. 1575-1580, 1986.
- [12] E. Peli, "Visual issues in the use of a head mounted monocular display," *Optical Engineering*, vol. 29, pp. 883-892, 1990.
- [13] E. Peli, "Simple 1-D enhancement for head-mounted low vision aid," *Visual Impairment Research*, vol. 1, pp. 3-10, 1999.

- [14] E. Peli, J. Kim, Y. Yitzhaky, R. B. Goldstein, and R. L. Woods, "Wideband enhancement of television images for people with visual impairment," *Journal of the Optical Society of America A*, vol. 21, pp. 937-950, 2004.
- [15] E. Peli, "Feature detection algorithm based on a visual system model," *Proceedings of the IEEE*, vol. 90, pp. 78-93, 2002.
- [16] G. Luo, N. M. Rensing, E. Weststrate, and E. Peli, "Registration of an on-axis see-through head mounted display and camera system," *Optical Engineering*, vol. 44, 2005.
- [17] J. Tang, E. Peli, and S. Acton, "Image enhancement using a contrast measure in the compressed domain," *IEEE Signal Processing Letters*, vol. 10, pp. 289-292, 2003.
- [18] J. Tang, J. Kim, and E. Peli, "Image enhancement in the JPEG domain for people with vision impairment," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 2013-2023, 2004.
- [19] J. Kim, A. Vora, and E. Peli, "MPEG-based image enhancement for the visually impaired," *Optical Engineering*, vol. 43, pp. 1318-1328, 2004.
- [20] PixelTools, "MPEG Repair software, Pixeltools Corporation, <http://www.pixeltools.com/MpegRepair.html>," 2006.
- [21] M. Fullerton and E. Peli, "Post-transmission digital video enhancement for people with visual impairments," *Journal of the Society for Information Display*, vol. 14, pp. 15-24, 2006.
- [22] D. Miller and G. Benedek, *Intraocular light scattering*. Springfield, Illinois: Charles C. Thomas, 1973.
- [23] E. Peli, "Contrast in complex images," *Journal of the Optical Society America A*, vol. 7, pp. 2032-2040, 1990.
- [24] E. Peli, "Test of a model of foveal vision by using simulations," *Journal of the Optical Society of America A*, vol. 13, pp. 1131-1138, 1996.
- [25] E. Peli, "Contrast sensitivity function and image discrimination," *Journal of the Optical Society of America A*, vol. 18, pp. 283-293, 2001.
- [26] E. Peli, "Enhancement of retinal images: pros and problems," *Neuroscience and Behavioral Reviews*, vol. 17, pp. 477-482, 1993.
- [27] J. Sergent, "Microgenesis of face perception," in *Aspects of Face Processing*, H. Ellis, M. Jeeves, F. Newcombe, and A. Young, Eds. Boston: Martinus Nijhoff, 1986, pp. 17-33.
- [28] A. O'Toole, R. Millward, and J. Anderson, "A physical system approach to recognition memory for spatially transformed faces," *Neural Networks*, vol. 1, pp. 179-199, 1988.
- [29] C. E. Metz, P. Wang, and H. B. Kronman, "A new approach for testing the significance of differences between ROC curves measured from correlated data," presented at Proceedings of the VIII Conference on Information Processing in Medical Imaging, The Hague, 1983.
- [30] J. A. Swets and R. M. Pickett, "Fundamentals of accuracy analysis," in *Evaluation of Diagnostic Systems: Methods From Signal Detection Theory*. New York: Academic Press, 1982.
- [31] E. Peli, E. M. Fine, and A. T. Labianca, "Evaluating visual information provided by audio description," *Journal of Visual Impairment & Blindness*, vol. 90, pp. 378-385, 1996.
- [32] B. J. Cronin and S. R. King, "The development of the descriptive video service," *Journal of Visual Impairment and Blindness*, vol. 84, pp. 503-506, 1990.
- [33] E. Fine, E. Peli, and N. Brady, "Video enhancement improves performance of persons with moderate visual loss," presented at Proceedings of the International Conference on Low Vision, "Vision '96", Madrid, Spain, 1996.
- [34] R. Hamberg and H. de Ridder, "Continuous assessment of perceptual image quality.," *Journal of the Optical Society of America A*, vol. 12, pp. 2573-2577., 1995.
- [35] M. Fullerton and E. Peli, "People with visual impairment prefer TV viewing using a contrast enhancement consumer product," presented at 2007 SID International Symposium, 2007 (Submitted).