MPEG-based image enhancement for the visually impaired

Jeonghoon Kim
Harvard Medical School
The Schepens Eye Research Institute
20 Staniford Street
Boston, Massachusetts 02114-2500
and
Shinheung College
Department of Electronic Communication
#117, Howon dong
Uijeongbu-City
Gyeonggi-do, 480-701, Korea

Avni Vora
Eli Peli, MEMBER SPIE
Harvard Medical School
The Schepens Eye Research Institute
20 Staniford Street
Boston, Massachusetts 02114-2500
E-mail: eli@vision.eri.harvard.edu

Abstract. An MPEG-based image contrast enhancement algorithm for people with low vision is presented. Contrast enhancement is achieved by modifying the inter- and intra-quantization matrices in the MPEG decoder during the decompression stage. The algorithm has low computational complexity and does not affect the MPEG compressibility of the original image. We propose an enhancement filter based on the visual characteristics of low-vision patients, and report the results of preference experiments with 24 visually impaired subjects. Subjects favored low to moderate levels of enhancement for two of the tested video sequences, but favored only low levels of enhancement and rejected higher enhancement for two other sequences that had fast motion. [DOI: 10.1117/1.1723493]

Subject terms: television enhancement; MPEG decoding; spatial frequency; visual acuity; central field loss; motion compensation.

Paper 030364 received Jul. 28, 2003; revised manuscript received Dec. 13, 2003; accepted for publication Dec. 21, 2003.

1 Introduction

As the population ages, a growing number of people suffer from visual impairments. These impairments and their resulting disabilities greatly impact the quality of life of many elderly people. A Louis Harris survey found that vision impairment affects 17% of Americans age 45 and older, and 26% of those age 75 and older.[1]

Traditionally, vision rehabilitation research was aimed at improving mobility and reading abilities. But more recently, efforts have been made to improve the ability of the visually impaired to perceive pictorial information with the use of video-based devices. Magnification partially compensates for the loss of sensitivity experienced by the visually impaired and improves their ability to perceive visual information. But magnification was shown to be insufficient to restore some functions such as reading rate and face recognition.[2] Contrast enhancement of video may improve the visibility further and thus help the visually impaired enjoy television.

Digital image enhancement to improve video images for the visually impaired was first proposed by Peli and Peli[3] (applying an adaptive enhancement algorithm), and Peli, Arend, and Timberlake[4] investigated the use of a number of common image enhancement algorithms. Similar techniques were applied to the enhancement of text by Lawton[5] and by Fine and Peli.[6] While image enhancement was shown to modestly improve reading rate and may improve mobility as well, we see the main value of image enhancement in providing the visually impaired with better access to the ubiquitous video images presented on monitors. Television is an important means of obtaining information and sharing in the culture. Since television is primarily a visual medium, visually impaired people have not had full access to its benefits. Even so, most people with visual impairments do watch television with their families, and prefer watching television to other activities.[7,8] Television use by the visually impaired has increased over the years, and visually impaired people watch television nearly as much as or more than normally sighted people.[7,9] The Descriptive Video Service (DVS), which broadcasts programs with a separate audio channel carrying a narrative description of the visual scene,[9] is available for the visually impaired. Although it is helpful (particularly for the blind), it remains a limited service, and substituting auditory for visual information can detract from the television-viewing experience.

Enhancement of static images was shown to significantly and substantially improve face recognition for people with central visual field loss and for those with optical media opacities.[10] Real-time processing of live color video, using an adaptive enhancement algorithm,[3] was made possible with the development of the DigiVision device.[11] A pilot study using this device found increased recognition of details in the videos, and almost 95% of the subjects preferred the individually tuned enhancement this device offered over unenhanced videos.[6] A different study, using a face recognition task (in static images), found that individually tuned enhancement improved recognition, but no better than uniformly applied adaptive enhancement.[12] An additional study of motion video found that subjects significantly preferred enhanced images to unenhanced images, and that individual selection of parameters resulted in a greater affinity for enhanced video over unmodified video.[13]

Most of the previous approaches were based on the filtering of analog (uncompressed) video, even though digital signal processing was used. However, the use of digital video products applying MPEG compression (see Sec. 2.1 for an MPEG primer), such as digital televisions, DVD players, and digital camcorders, is rapidly growing. The global sales of DVD players were estimated to reach about 41 million units in 2002 and 52 million units in 2003.[14] The

Federal Communications Commission (FCC) has adopted a plan requiring digital television (DTV) tuners on nearly all new television sets,[15] giving consumers access to digital programming. Techniques for video enhancement to aid visually impaired people must evolve to be compatible with new digital media formats.

There are various approaches to enhancing digital compressed images. Images may be enhanced prior to compression and coding (precompression), after decoding (postdecompression), or within the coding/decoding process (which is the method we propose here). Precompression enhancement may affect the compressibility of the image and may require postdecompression processing to maintain quality.[16,17] The postdecompression approach[18–22] can be adopted without affecting the compressibility of the original image. But it tends to increase the severity of compression artifacts (e.g., block artifacts), making them more clearly visible.[18,19,23]

Some postdecompression enhancement algorithms use coding information from compressed video to improve performance.[20–22] For example, Tsai et al.[20] have proposed an iterative algorithm using coding information for enhancing video sequences that are encoded at very low bit rates. The algorithm improved video quality in peak signal-to-noise ratio (PSNR) gain, but was limited by high computational complexity. Boroczky and Janssen[21] proposed deriving a usefulness metric for enhancement (UME) using compression information from MPEG-2 bitstreams to improve the performance of sharpness enhancement. Yang and Boroczky[22] further improved on the original concept[21] by redefining the UME, as a quantitative value describing whether and how much a pixel could be enhanced without worsening coding artifacts. The new UME algorithm has lower computational complexity than the Tsai et al. iterative algorithm,[20] but it is still computationally expensive.

Various postprocessing methods have recently been developed to remove block artifacts that occur after image decompression.[17,23–28] While reduction of compression artifacts in still images has been studied extensively, little work has been done in improving the quality of compressed video.[23] Martucci[28] removed block artifacts in the compressed frequency domain, but the process required modification of the standard discrete cosine transform (DCT) configuration in JPEG or MPEG. Konstantinides and Beretta[29] implemented an image sharpening technique in the JPEG domain using the quantization matrix in the decoding stage, similar to our own method.[30,31] They took a degraded image (as produced by imaging systems, such as color scanners or fax machines) and used a reference image to attempt recovery of the original image quality. The sharpening algorithm showed promising results, but was applied only to static images where a high-quality reference image was available.

To compensate for their reduced resolution and contrast sensitivity, low-vision patients tend to view television monitors at very close distances. From such a short distance, they are much more sensitive to block artifacts from image compression than are most television viewers. Thus, minimizing the appearance of block artifacts is an important consideration when contrast-enhancing compressed images for low-vision patients.

We developed an MPEG-based video enhancement that operates during the decompression stage. This approach can reduce the appearance of block artifacts and is based on the visual characteristics of low-vision patients.

Our MPEG enhancement algorithm is based on an algorithm for the enhancement of JPEG compressed images.[30,31] In the previous JPEG enhancement, the filtering was applied to all DCT frequencies without considering the visual properties or viewing distances typical of people with low vision. The MPEG-based enhancement approach presented here is constrained by two considerations. The enhancement approach must be compatible with the close viewing and the specific contrast sensitivity of people with low vision, and it must also interoperate with the current MPEG-2 standard that handles DTV. Because the enhancement is performed simultaneously with decompression and it only requires access to the quantization matrix, it has minimal computational cost.

2 Image Enhancement in the MPEG-2 Domain

2.1 *MPEG Basics*

The international standard ISO/IEC 13818-2 ‘‘Generic coding of moving pictures and associated audio information: video,’’ and the ATSC document A/54 (Ref. 32) ‘‘Guide to the use of the ATSC digital television standard,’’ describe a system known as MPEG-2 for encoding and decoding digital video data. An MPEG system is composed of an encoder and a decoder.[33–35] The basic underlying process is similar to JPEG coding. Encoded pictures are made up of pixels. Each nonoverlapping array of 8×8 pixels is known as a block. A 2×2 array of blocks (16×16 array of pixels) is termed a macroblock. The 2-D DCT is computed for each block. The DCT coefficients are quantized by dividing each coefficient by the corresponding entry in a quantization matrix and then rounding to integers. Quantization of the DCT coefficients is a lossy compression process. Many small coefficients are quantized to zero in this step. A zig-zag scan of the DCT matrix and (lossless) entropy coding take advantage of these zero-valued entries to lower the bit rate required to encode the coefficients for storage or transmission. These compression steps combined take advantage of spatial correlations in the image. The quantization matrix is transmitted with the encoded image in the header that contains, in addition to the quantization matrix, other encoding and displaying parameters such as the size of the coded video pictures, image aspect ratio, picture rate, etc.[35]

MPEG compression also takes advantage of the temporal correlation between frames. Compression is achieved using prediction techniques (motion estimation in the encoder, motion compensation in the decoder). Some video frames called ‘‘I’’ (intra) pictures are encoded independently of all other pictures without prediction. Pictures called ‘‘P’’ (predicted) pictures may be encoded using prediction from previous pictures (which may be intra- or predicted pictures) and will in general be used as a reference for other predicted pictures. ‘‘B’’ (bidirectional predicted) pictures may be encoded using prediction from both previous and subsequent coded pictures and provide the highest amount of compression. The P and B pictures are called inter-frames. Each macroblock in these frames is classified as either moving or nonmoving. Nonmoving blocks of the inter-frames contain essentially the same information as the blocks in the preceding picture(s), and thus do not have to be recoded or retransmited. In the case of the inter-frame moving blocks, a coding process similar to that used for intra-coded blocks is applied using the inter-quantization matrix, which may be different from the intra-quantization matrix.

Macroblocks from a P or B picture to be encoded are fed to both a subtractor and a motion estimator. The motion estimator compares each of these new macroblocks with macroblocks in a previously stored reference picture or pictures. It finds the macroblock in the reference picture that most closely matches the new macroblock. The motion estimator then calculates a motion vector, which represents the horizontal and vertical displacement from the macroblock being encoded to the matching macroblock-sized area in the reference picture.

The motion estimator also reads this matching macroblock (known as a predicted macroblock) out of the reference picture memory and sends it to the subtractor that subtracts it, on a pixel by pixel basis, from the new macroblock entering the encoder. This forms an error prediction or residual block that represents the difference between the predicted macroblock and the actual macroblock being encoded. This residual is often very small, permitting a higher level of compression. The residual block is coded using a DCT and quantization. However, a different quantization matrix, the inter-quantization matrix, is applied to the residual or moving sections in the video sequence, while the intra-quantization matrix is applied to static nonmoving blocks of the intra-pictures in the video sequence.

In the MPEG decoder, the video sequence is decompressed.[35] The decoding process can be thought of as the reverse of the encoding process. The received encoded data is first losslessly decoded. The quantization matrices are derived from the header. Quantized DCT coefficients are fed to the inverse quantizer and then to an inverse DCT that transforms them back to the spatial domain. For P and B pictures, the motion vector data are used to read a particular macroblock (a predicted macroblock) out of a previously stored reference picture. Adding this prediction to the residual forms the reconstructed picture data. In the case of a motion block, the residual is decoded using the inter-quantization matrix. For I pictures, there are no motion vectors and no reference picture, so the prediction is forced to zero and the intra-quantization matrix is used. For I and P pictures, the output is fed back to be stored as a reference picture for future predictions.

2.2 *Spatial Frequency Filtering in DCT Domain*

The properties of the DCT coefficients provide a very natural way for defining spatial frequency filters in the DCT domain.[28] Figure 1 is an illustration of DCT basis functions for 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 content. Down the left column, the functions increase in vertical spatial frequency, with an increase in both horizontal and vertical frequencies along the diagonals. The normalized spatial frequency, f_n (cycles/pixel) of the corresponding basis functions in the DCT domain, is

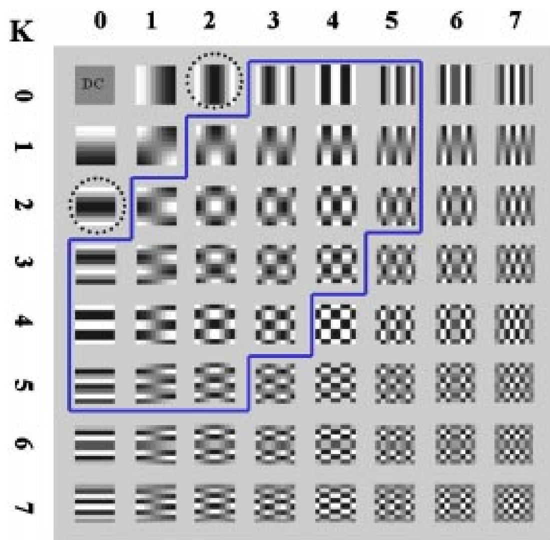

Fig. 1 The DCT basis functions for an 8×8 block. The coefficients represent the magnitude of the basis function representing the signal. *K* is the order of the basis function and the corresponding coefficients derived from the DCT. As can be seen, the basis functions order *K* represent the spatial frequency of *K*/2 cycles/block. The basis functions inside the lined area represent the critical frequencies to be enhanced. We excluded the two functions circled because their enhancement increased the appearance of block artifacts. In areas outside these bands, the quantization matrix was not modified.

$$f_n = K/2N, \quad K = 0,1,2,...N-1, \tag{1}$$

where K is the order of the coefficient and N is the size of the block.

Effective image enhancement requires increasing the contrast in a specific range of frequencies. Increasing the contrast at spatial frequencies that are not at all visible provides no benefit, while increasing the contrast of already visible frequencies can cause distortions and is also unlikely to provide any benefit. Enhancement can be effective, however, at frequencies to which the viewer is sensitive only to high contrast levels. Figure 2 shows the contrast detection threshold as a function of spatial frequency adopted from Peli et al.[10] Those low-vision patients [visual acuity 0.48 to 0.83 log minimum angle of resolution (MAR)] could detect test targets in the spatial frequency range of 3 to 7 cycles/deg, but required much higher contrast than normally sighted observers. Most of them could not detect at all targets at frequencies higher than 8 cycles/deg. To relate the spatial frequencies to orders in the DCT domain and to the low-vision patient's contrast detection thresholds, the spatial frequency variable f in cycles/deg of the contrast threshold function is converted to the normalized spatial frequency f_n in cycles/pixel as follows.[36]

$$f_n(\text{cycles/pixel}) = f(\text{cycles/deg})/f_s(\text{pixels/deg}), \tag{2}$$

where the sampling frequency (f_s) depends on the viewing distance and the screen size.

In a previous study in our lab, the median preferred viewing distance of low-vision subjects for watching a 27-in. television screen (720×480 pixels) was found to be 36 in. The f_s for this distance is approximately 22 (pixels/deg).

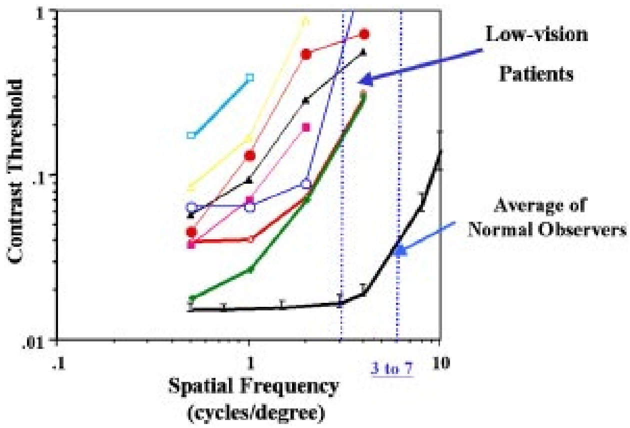

Fig. 2 The contrast detection threshold as a function of spatial frequency for low-vision patients and normally sighted observers (modified from Peli et al.[10]). The bolded black line shows the average contrast detection threshold of people with normal vision. The thinner lines show the contrast thresholds of eight low-vision patients. The range of 3 to 7 cycles/deg is the range of frequencies where enhancement has the most potential benefit for low-vision patients. Note that patients could not see the frequencies higher than 8 cycles/deg.

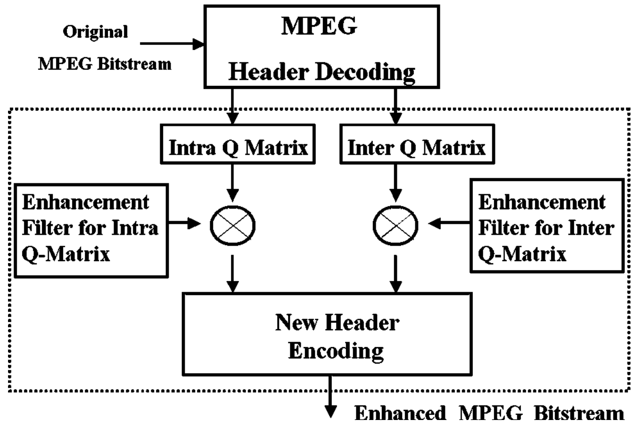

Fig. 3 Data flow diagram for image enhancement in the MPEG decompression domain. Still and motion blocks can be enhanced by filtering of intra- and inter-**Q** matrices, respectively. Note that ⊗ is a point-by-point multiplication. The header decoding and new header encoding were implemented in software.[41]

By substituting $K=7$, $N=8$, and $f_s=22$ into Eqs. (1) and (2), the maximum visual spatial frequency in the 8×8 block is 9.6 cycles/deg. The conversion results for these conditions are given in Table 1. In the DCT domain, we enhanced the shaded frequency orders of $K=2$ to $K=5$ to achieve enhancement of the visual frequency range of approximately 3 to 7 cycles/deg. In Fig. 1, the outlined area shows the basis functions we enhanced. We enhanced every component in the $K=2$ to 5 bands except the two circled DCT basis functions. Removing the enhancement from these low-frequency coefficients reduced the severity of block artifacts.

Table 1 Visual spatial frequencies corresponding to the DCT orders of basis functions for a viewing distance of 36 in. and a 27-in. television monitor.

K (DCT order)	f_n (cycles/pixel)	Visual frequency (cycles/deg)
7	0.44	9.6
6	0.38	8.3
5	0.31	6.8
4	0.25	5.5
3	0.19	4.1
2	0.13	2.8
1	0.06	1.4
0	0	0

2.3 *Image Enhancement Using Quantization Matrices*

Applying filtering in the DCT domain can be achieved in the MPEG decompression stage by manipulating the **Q** matrices available in the sequence header. As explained in Sec. 2.1, in MPEG, there are two different **Q** matrices—intra- and inter-matrices—with different values for quantization of still and moving macroblocks. The intra-**Q** matrix contains values for quantizing still macroblocks, while the inter-**Q** matrix operates on moving blocks. In our enhancement approach, both the intra- and inter-**Q** matrices are multiplied, point by point, with predesigned 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. The filtration is applied as:

$$\overline{q}_{ij}=ef_{ij}\cdot q_{ij}, \tag{3}$$

where q_{ij} are the elements of intra- or inter-quantization matrices, ef_{ij} are the elements of enhancement filters (EF), and $\overline{q}_{ij}$ are the elements of the modified intra- or inter-quantization tables **Q**, which are then used in the MPEG decoding. The MPEG header decoding and new header encoding operations, shown in Fig. 3, were implemented using ReStream™ (Siegfried Hildebrarnd, Deggendorf, Germany) software.[37] In the preference study reported here, the same filters were applied to both intra- and inter-matrics.

$$EF=\begin{pmatrix} 1 & 1 & 1 & \lambda & \lambda & \lambda & 1 & 1 \\ 1 & 1 & \lambda & \lambda & a\lambda & \lambda & 1 & 1 \\ 1 & \lambda & \lambda & \lambda & a\lambda & \lambda & 1 & 1 \\ \lambda & \lambda & \lambda & a\lambda & \lambda & 1 & 1 & 1 \\ \lambda & \lambda & \lambda & \lambda & 1 & 1 & 1 & 1 \\ \lambda & \lambda & \lambda & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}. \tag{4}$$

Equation (4) shows the enhancement filter (EF) array applied in our study. The lambda (λ) parameter is an enhancement gain that might be modified by the user in real time from a remote control of a device implementing our algorithm. Enhancement of interlaced television formats tends

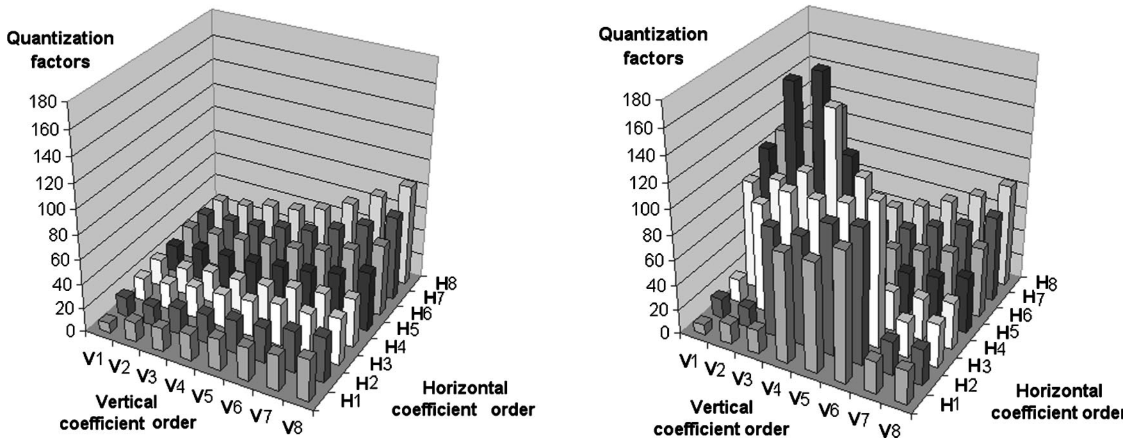

Fig. 4 A graphic illustration of the quantization factors used: (a) the default intra-**Q** matrix, (b) its filtered/enhanced **Q** matrix (using $\lambda=4$, $a=1.5$). The ratios between (b) and (a) represent the enhancement applied at different frequencies.

to increase line flickering artifacts. Previously, we used directional enhancement (vertical edge enhancement only) in an effort to reduce such flickering.[30] Here, we implemented a slight asymmetry with the placement of the arbitrary factor a in the filter structure, emphasizing vertical edge enhancement, as shown in Eq. (4). This was sufficient to substantially reduce these artifacts. Figure 4 illustrates the quantization factors used in the MPEG standard default intra-**Q** matrix[35] [Fig. 4(a)] and its filtered **Q** matrix [Fig. 4(b)] using Eqs. (3) and (4) with ($\lambda=4$, $a=1.5$).

3 Generation of Enhanced Test Video Sequences

Table 2 lists the four digital standard definition (SD)-grade MPEG-2 elementary test sequences (resolution: 704×480) that we used.[38,39] SD grade is frequently used in DVD and digital television systems in which the quality is approximately equivalent to that of current NTSC television.[32] The interlace scanned Susie, Flowers (Flwer), and Table Tennis (Tennis) sequences are available at seven different bit rates (40, 18, 12, 8, 6, 4, and 1.5 Mbps). We chose the 8 Mbps, as it is the median bit rate. For the Lion sequence, we used a lower bit rate of 6 Mbps because the Lion sequence is a progressive sequence and thus has slightly higher quality. Generally, rates of 6 to 8 Mbps are used to fit 90 min of video onto a 4.7-GB DVD disk.[40]

We carried out pilot tests with nine visually impaired subjects to decide on a number of experimental parameters:

- range and step size of enhancement levels
- mode of presentation—sequential versus side by side (split screen)
- sequence duration
- number of sequences and levels to be tested and repeated.

Table 2 The characteristics of the MPEG-2 sequences used.

Name	Lion.m2v	Susie.m2v	Flwer.m2v	Tennis.m2v
Bit rate	6 Mbps	8 Mbps	8 Mbps	8 Mbps
Scan	Progressive	Interlaced	Interlaced	Interlaced

Based on the results of these pilot tests, we created enhanced MPEG-2 video sequences using a range of λ values ($\lambda=2$, 3, 4, and 5) with a constant factor $a=1.5$, for both the intra- and inter-matrices (see Appendix A in Sec. 7). For comparison of original and enhanced sequences, we used a side-by-side (split-screen) display as shown in Figs. 5–8. To create the side-by-side display, we decoded original and enhanced MPEG sequences with the MPEG software decoder.[41] After decoding the MPEG sequences (including the required enhancement), we cut each sequence so that it was only half the original width, but maintained the center of the picture. We then merged the original and enhanced sequences using Matlab (MatLab Image Processing Toolbox, MathWorks, Natick, Massachusetts) so that they played simultaneously. We mirror reversed the placement

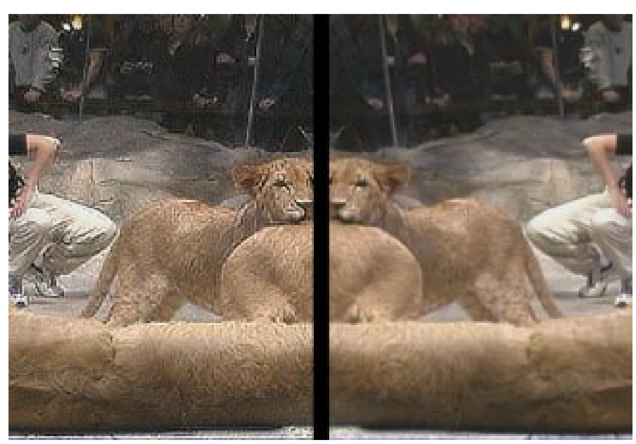

Fig. 5 An example of side-by-side (split) screen view of the Lion sequence used in the experiment. Here the left side is an enhanced video ($\lambda=4$) and the right side is an original video in mirror image. Only half the widths of the original and enhanced (352 pixel) videos were used. The two halves were merged into one video sequence. Most subjects favored a moderate level of enhancement for this sequence.

Fig. 6 An example of the Susie sequence ($\lambda=4$). This sequence is an interlaced sequence with moderate motion. Most subjects favored a moderate level of enhancement for this sequence.

Fig. 8 An example of the Flwer sequence ($\lambda=4$). This sequence is an interlaced sequence with fast motion. Most subjects favored little enhancement for this sequence.

of the half video to enable the side-by-side comparison of similar image areas. Both left- 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$\times$4 gains$\times$2 sides) each 5-sec in length were generated this way. Presentation of the 32 sequences took about 30 min. Experiments longer than 30 min. might have been too fatiguing for our mostly elderly subjects.

4 Experimental Evaluation

4.1 *Subjects*

24 visually impaired subjects (14 men and 10 women), ranging in age from 44.8 to 85.7 years (median 71.0 years) participated in the study. The subjects' visual acuity was measured using an electronic test device with random letter presentation (BVAT model number 22-4850, Mentor O&O Incorporated, Norwood, Massachusetts). Log MAR acuity ranged from 0.54 (20/70) to 2.10 (20/2500) (average 1.02 $\pm$0.35). All subjects had documented central field loss in both eyes. Visual field was measured using an auto-plot tangent screen (Cat. Number 71-54-41, Bausch and Lomb, Rochester, New York). The fields were measured monocularly, using a 6-mm white target at a distance of one meter. The central blind spots ranged form 7 to 50 deg in the horizontal dimension (median 20 deg) and from 10 to 50 deg in the vertical dimension (median 30 deg). The average time for measuring visual acuity and visual field was about 45 min.

Fig. 7 An example of the ''Table Tennis'' sequence ($\lambda=4$). This sequence is an interlace sequence with fast motion. Most subjects favored little enhancement for this sequence.

4.2 *Procedures*

Subjects were asked to sit approximately 36 in. from the 19-in. monitor (Dell Dimension 8250 computer with a Dell P1130 Color Monitor, Resolution 1600$\times$1200, 32-bit color at a 75-Hz refresh rate, graphic card NVIDIA Gforce4 MX420, luminance range 4 to 104 cd/m^2) in a dimly lit room (3.6 foot candles). They were shown the 5-sec video sequences, which repeated until the subject responded. The subjects were asked to evaluate each side of the video sequence for how clear the video was, how much detail and information could be obtained from the video, and the general quality of the picture. Using these guidelines, they were asked to choose which side of the video (left or right) they preferred. If the subject could not see any difference in the two sides at the first test sequence (levels 4 or 5), they were allowed to move closer to whatever distance they chose. Measurements of their viewing distances were taken. Subjects were forced to choose a side (i.e., they could not say the pictures looked the same). 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'' (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 unenhanced 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.

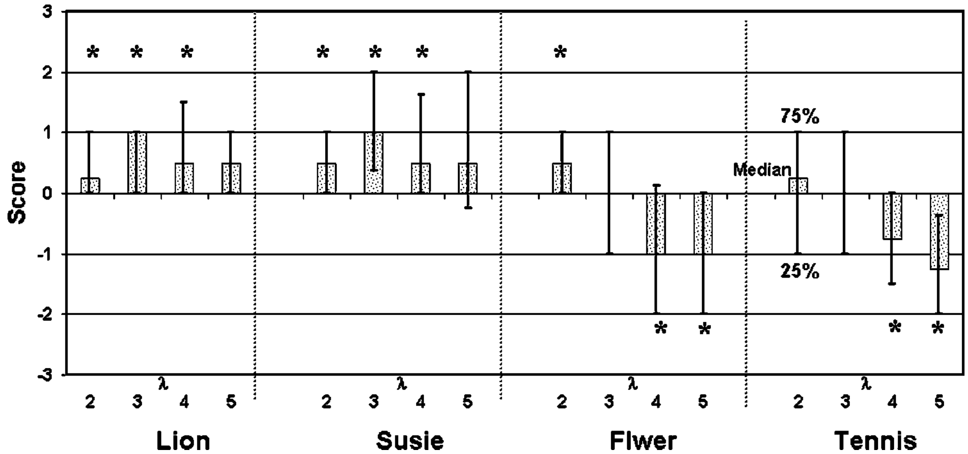

Fig. 9 The median (50%) scores of the 24 subjects' responses to the different sequences and levels of enhancement. The error bars show the range from first quartile (25%) to third quartile (75%). The subjects preferred moderate levels of enhancement for the Lion and Susie sequences. For the two other sequences, the subjects only favored a low level of enhancement ($\lambda=2$). Note that the * indicate a statistically significant effect at the $p<0.05$ level.

5 Results

Figures 5–8 are examples of the test images from the sequences presented to the subjects. In all figures, the left side is a frame from the enhanced sequence ($\lambda=4$) and the right side is the corresponding frame from the original sequence. The subjects as a group preferred the three lower enhancement levels ($\lambda=2$ to 4) for two of the sequences: Susie (Wilcoxon signed rank test, $Z_{23}>2.26$, $p<0.03$) and Lion ($Z_{23}>2.55$, $p<0.02$) (Fig. 9). The small preference shown for the highest enhancement level ($\lambda=5$) only approached significance ($p=0.08$ and $p=0.07$ for Lion and Susie, respectively). The two highest enhancement levels ($\lambda=4$ and 5) for the two other sequences (Flwer and Tennis) were rejected ($Z_{23}>2.48$, $p<0.03$; and $Z_{23}>2.36$, $p<0.02$, respectively). The lower enhancement levels were not significantly different from the original, although there was a slight preference for the low level of enhancement ($\lambda=2$), and that effect was statistically significant for the Flwer sequence ($Z_{23}=2.14$, $p=0.03$).

During the experiments, we noted that a few subjects seemed to have a clear preference for one side of the screen irrespective of the presentation of enhanced or original sequence. In each condition, there were two identical presentations, one with the enhanced sequence on the right and one with the enhanced sequence on the left. We therefore tested each patient to see if the selection was the same for the two presentations or different. 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.

Figure 10 shows the results from the 13 subjects who showed consistent (unbiased) preference regardless of the position of enhanced sequences (i.e., on the left or right side of the screen). The results of these 13 subjects are similar to those of the whole group. There is slightly higher preferences for the enhancement of the Lion, from $\lambda=2$ to 4 ($Z_{12}>2.05$, $p<0.05$), and of Susie, from $\lambda=3$ to 4 ($Z_{12}>2.15$, $p<0.04$). For the other two sequences, subjects significantly rejected the two highest levels ($\lambda=4$ to 5)

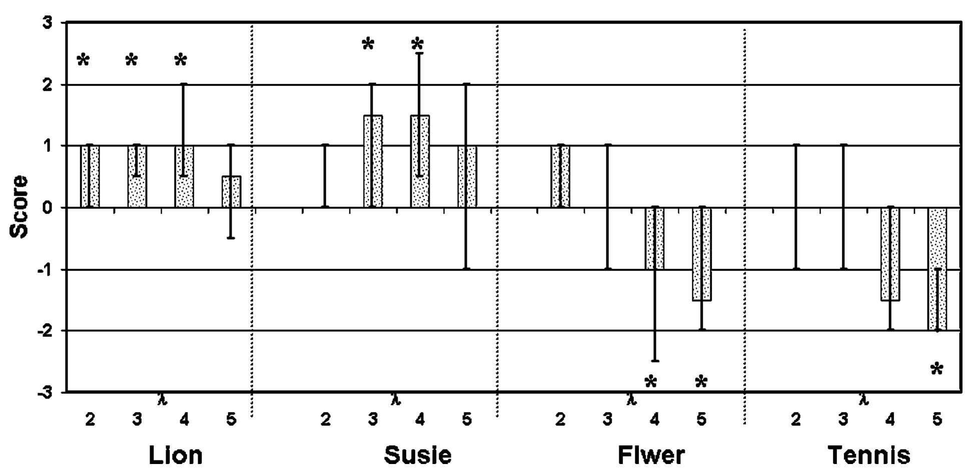

Fig. 10 The median values of the 13 subjects who did not have a bias to one side or another. The 13 subjects' results show similar tendencies as the whole group of 24 subjects shown in Fig. 9.

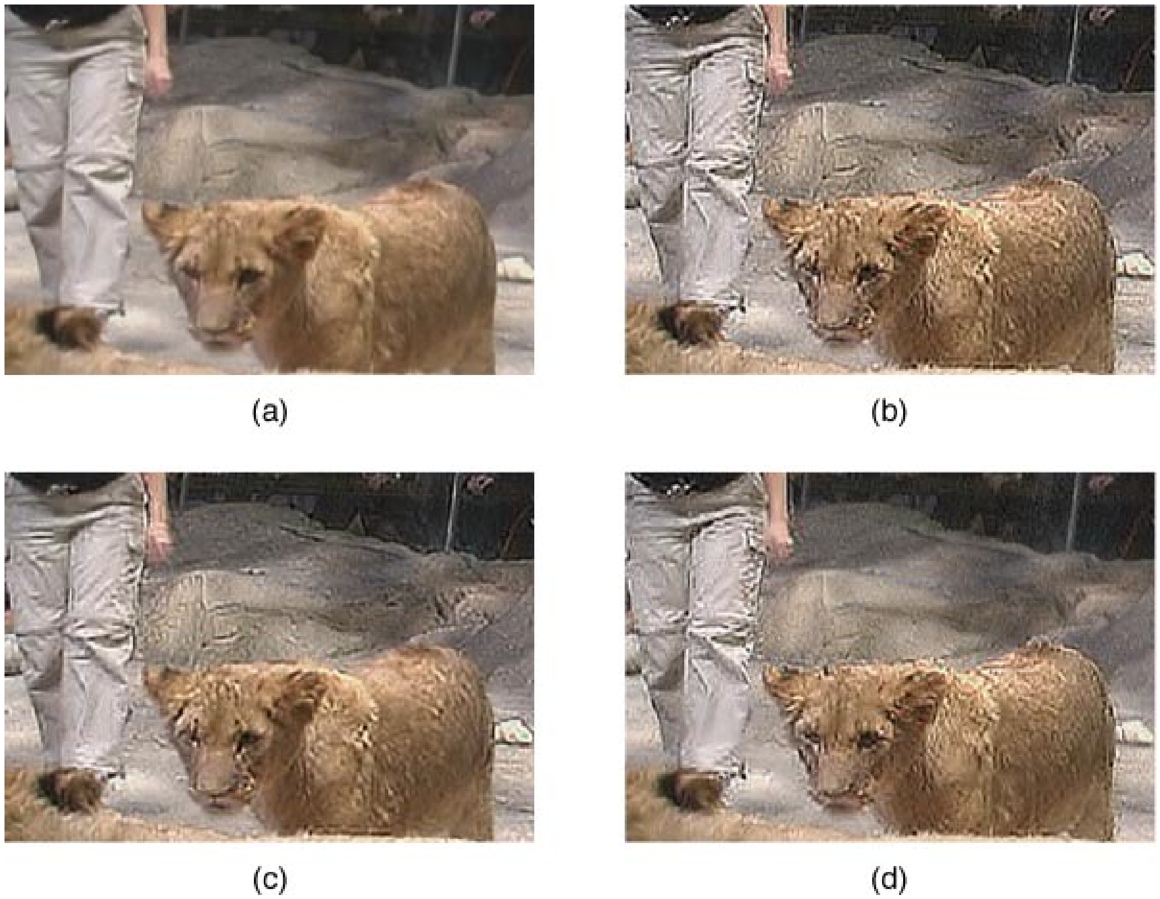

Fig. 11 The effects of intra- plus inter-, just intra-, and just inter-enhancement with $\lambda=4$ (the preferred enhancement level used in the Lion sequence enhancement). (a) The original image is an unenhanced B (bidirectional) picture. (b) The intra- plus inter-enhancement shows combined enhancing effects in moving and still areas. We used this intra- and inter-enhancement for our experiment. (c) The intra-enhancement shows the enhancement effects in all still areas. (d) The inter-enhancement shows the enhancement effects in the moving lion and the person's trousers.

for Flwer ($Z_{12}>2.27$, $p<0.03$), and the highest level ($\lambda=5$) for Tennis ($Z_{12}>2.76$, $p<0.01$).

The median preferred viewing distance was found to be 20 in. (minimum of 8 in., maximum of 32 in.) from the 19-in. PC monitor. From 20 in., the presented images spanned 42 and 29 deg, horizontally and vertically, respectively. The f_s for this distance was approximately 16 (pixels/deg), and the maximum visual spatial frequency was 7 cycles/deg for the 8×8 block. Thus, in our case, the frequency bands of $K=2$ to $K=5$ we enhanced corresponded to approximately 2 to 5 cycles/deg.

6 Conclusions

We implemented and tested a new MPEG-based television image enhancement for people with low vision. The enhancement is applied during the MPEG decompression phase and only requires access to the quantization matrices. As such, the computation load is minimal and can be easily applied in real time. 13 of the 24 visually impaired subjects, who had no side bias in their response, favored a moderate level of enhancement for the Lion and Susie sequences, which are representative of the amount and type of camera motion in popular television program formats such as drama or news. Subjects favored only low-level enhancement for the Flwer and Tennis sequences, and clearly rejected the higher levels of enhancement for these two sequences. It is possible that the enhancement for these sequences was rejected because these sequences contain more motion, and the enhancement of fast motion sequences resulted in visible motion artifacts or led to stronger enhancement artifacts due to the combined enhancing effects of the intra- and inter-enhancement (see Appendix A in Sec. 7).

We intend to investigate a few ways to reduce these motion artifacts in future studies. Reducing the enhancement level of interenhancement relative to the intraenhancement [e.g., intraenhancement level $(\lambda)=4$, interenhancement level $(\lambda)=2$ or 3] should result in the reduction of motion artifacts. Further, the information about motion is available within MPEG video coding and could be used to adjust the enhancement levels adaptively for video segments depicting fast motion or simply for all motion blocks. Others[21,22] have previously used motion estimation and scene change detection to ensure temporal consistency and control the gain of enhancement of MPEG video. Such

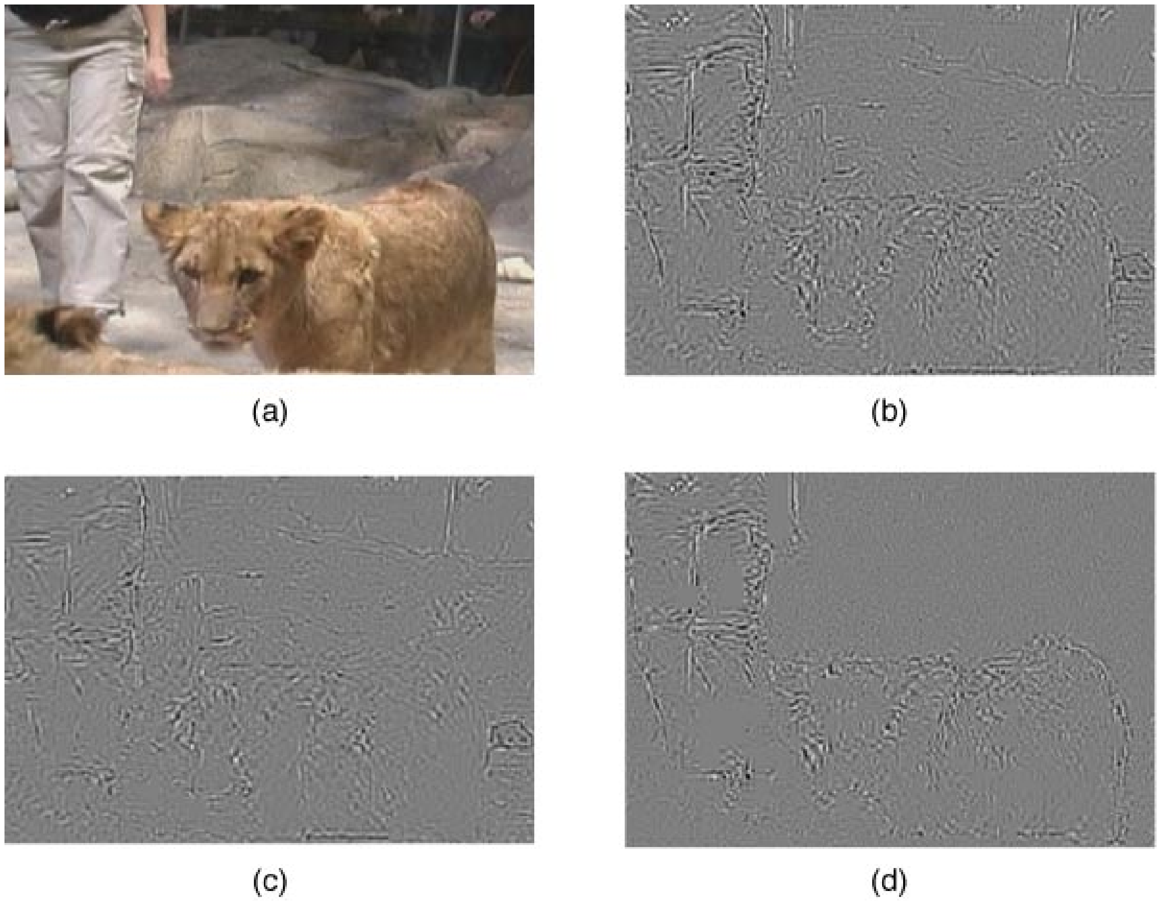

Fig. 12 The differences between (a) the original and each enhanced image [Figs. 11(b), 11(c), and 11(d)]. (b) The intra- plus inter-enhancement shows strong combined enhancement. (c) The intra-enhancement shows enhancement in all areas. (d) The inter-enhancement shows the enhancement effects in moving areas only. The static background in (d) is not enhanced at all.

adaptive gain control based on motion may provide even better results.

This MPEG-based image enhancement algorithm may provide an inexpensive and flexible way of delivering more visible digital video to elderly and visually impaired audiences. Implementation requires only minimal modification to conventional MPEG decoders. The enhancement could be individually tuned, and the level of enhancement could be varied in real time by the viewer using a remote control. This technology may have a wide market appeal for many elderly television and PC viewers with moderate visual impairments.

7 Appendix A: Effects of Enhancement During Decompression from Intra- and Inter-Matrix

MPEG coding and decoding apply different quantization matrices to different frames [I (intra), P (predictive), and B (bidirectionally predictive) pictures]. I pictures use the intra-**Q** matrix for all blocks. P and B pictures apply intra- and inter-**Q** matrices to still and moving blocks, respectively.[35] While the intra-matrix is used to quantize actual image blocks, the inter-matrix is used to quantize the image difference of moving blocks. The enhancement can be applied to either matrix alone or to both. When both matrices are modified, the modification can be identical or different.

7.1 *Intra-Only Enhancement*

Intra-only enhancement is enhancing intra-macroblocks with the intra-**Q** matrix. Only intra-macroblocks are used in I pictures and intra-macroblocks are used also in static segments of P and B pictures. Thus, the intra-only enhancement is enhancing still images and nonmoving blocks of images with motion.

$$I \rightarrow I_i',$$
$$P \rightarrow I_p' + \Delta p, \tag{5}$$
$$B \rightarrow \frac{I_i' + (I_p' + \Delta p)}{2} + \Delta b.$$

Equation (5) shows the effects of enhancing only the intra-matrix. If the current picture is an I picture, and the next picture is a P picture, the decompressed I picture will be enhanced only from intra-macroblocks with the enhancement filtering applied to the intra-**Q** matrix. The P picture will have some enhancing effects from the macroblocks forwarded from the previous I picture. However, the motion difference (Δp) blocks will be unmodified without inter-matrix enhancement. I_i' is an enhanced I picture from the intra-macroblocks and I_p' is a P picture partially enhanced with static forwarded macroblocks from the previous I picture. The B pictures may have enhancing effects from forward, backward, and bidirectional macroblocks. The bidirectional averaged macroblocks are widely used and will be enhanced except for motion differences, as shown in Eq. (5).

7.2 *Inter-Only Enhancement*

Inter-only enhancement is enhancing motion difference blocks by filtering the inter-**Q** matrix. Inter-**Q** matrix is used for the forward, backward, and bidirectional macroblocks in P and B pictures. Thus, as shown in Eq. (6), the inter-only enhancement can enhance the moving areas of P and B pictures. The $\Delta p'$ and $\Delta b'$ are enhanced motion difference components in both the P and B pictures.

$$I \rightarrow I_i,$$

$$P \rightarrow I_p + \Delta p', \tag{6}$$

$$B \rightarrow \frac{I_i + (I_p + \Delta p')}{2} + \Delta b'.$$

7.3 *Combined Intra- and Inter-Enhancement*

The enhancement filtering of combined intra- and inter-**Q** matrices will make combined enhancing effects of intra- and inter-**Q** matrices as shown in Eq. (7). This enhancement will enhance all the macroblocks, so both still and moving areas are enhanced together. While this is good for enhancement of the still areas, this combined enhancement may create stronger enhancement levels in the moving areas. This is because moving areas will be enhanced twice, once in the I picture and then again as a motion block, resulting in a double application of the enhancement to these blocks.

$$I \rightarrow I_i',$$

$$P \rightarrow I_p' + \Delta p', \tag{7}$$

$$B \rightarrow \frac{I_i' + (I_p' + \Delta p')}{2} + \Delta b'.$$

Figures 11(b), 11(c), and 11(d) illustrate the enhancement effects for a single video frame when processed with intra- and inter-matrices (as used in this study), and with the intra- or inter-enhancement alone. These frames were captured from MPEG decoded/enhanced videos. Figure 11(a) is an original B (bidirectional) picture decoded without enhancement. Figure 11(b) is the same picture enhanced with both intra- and inter-enhancement. Figure 11(c) is the picture enhanced only with intra-enhancement. Figure 11(d) shows the enhancing effect in the moving area only resulting from modifying the inter-matrix.

Figures 12(b), 12(c), and 12(d) illustrate the effects of each enhancement by presenting the differences between the enhanced frames [Figs. 11(b), 11(c), and 11(d)] and the original frame [Fig. 12(a)], respectively. It is evident that the intra-matrix enhancement [Fig. 12(c)] enhances the whole image, while the inter-matrix enhancement [Fig. 12(d)] results only in the enhancement of moving portions of the scene. The combined intra- and inter-enhancement in Fig. 12(b) thus has a stronger enhancement effect in the moving portions of the scene. We applied the combined enhancement here under the assumption that the motion blur that results in reduced sensitivity to moving patterns[42] would require stronger enhancement for such areas.

Acknowledgments

This work was supported in part by NIH grants EY05957 and EY12890 to Peli, and a postdoctoral fellowship program from Korea Science and Engineering Foundation (KOSEF) to Kim. We thank Robert B. Goldstein for experimental programming help and Russell L. Woods for advice regarding data analysis.

References

1. *The Lighthouse National Survey on Vision Loss: The Experience Attitudes, and Knowledge of Middle-Aged and Older Americans*, L. Harris and Associates. Inc., The Lighthouse, Inc. (New York, New York, 1995).
2. M. A. Bullimore, I. L. Bailey, and R. T. Wacker, "Face recognition in age-related maculopathy," *Invest. Ophthamol. Vis. Sci.* **32**, 2020–2029 (1991).
3. E. Peli and T. Peli, "Image enhancement for the visually impaired," *Opt. Eng.* **23**, 47–51 (1984).
4. E. Peli, L. E. Arend, and G. T. Timberlake, "Computerized image enhancement for low vision: New technology, new possibilities," *J. Vis. Impairement, Blindness* **80**, 849–854 (1986).
5. T. B. Lawton, "Improved reading performance using individualized compensation filters for observers with losses in central vision," *Ophthalmology* **96**, 115–126 (1989).
6. E. M. Fine and E. Peli, "Enhancement of text for the visually impaired," *J. Opt. Soc. Am. A* **12**, 1439–1447 (1995).
7. M. Berkowitz, L. G. Hiatt, P. de Toledo, J. Shapiro, and M. Lurie, *Characteristics, Activities and Needs of People with Limitation in Reading Print*, American Foundation for the Blind, New York (1979).
8. E. Josephson, *The Spatial Life of Blind People*, Vol. 19, American Foundation for the Blind, New York (1968).
9. B. J. Cronin and S. R. King, "The development of descriptive video service," *J. Vis. Impairment Blindness*, pp. 503–506 (1990).
10. 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 Ophthalmol. Vis. Sci.* **32**, 2337–2350 (1991).
11. 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," *SID 93*, pp. 491–494 (1993).
12. E. Peli, E. Lee, C. L. Trempe, and S. Buzney, "Image enhancement for the visually impaired: the effects of enhancement on face recognition," *J. Opt. Soc. Am. A* **11**(7), 1929–1939 (1994).
13. E. Peli, "Perceived quality of video enhanced for the visually impaired," in *Vision Science and Its Applications*, pp. 46–48, Santa Fe, NM (Optical Society of America, Washington, DC, 1999).
14. "Microsoft's DVD codec play stirs up consumer pot," *EE Times*, p. 14 (17 Dec. 2001).
15. "FCC introduces phase-in plan for DTV tuners," see http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-225221A1.doc?date=020808.
16. H. Akitsu and N. Kutsuwada, "Image quality improvement by pre and post processing in DCT coding," *Proc. SPIE* **1912**, 488–496 (1993).
17. O. Hader, A. Stern, and R. Koresh, "Enhancement of an image compression algorithm by pre- and post-filtering," *Opt. Eng.* **40**(20), 193–199 (2001).
18. H. Greenspan, C. H. Anderson, and S. Akber, "Image enhancement by nonlinear extrapolation in frequency space," *IEEE Trans. Image Process.* **9**(6), 1035–1047 (2000).
19. S. S. Agaian, "Transform-based image enhancement algorithms with

performance measure,'' *IEEE Trans. Image Process.* **10**(3), 367–382 (2001).
20. C. J. Tsai, P. Karunaratne, N. P. Galatsanos, and A. K. Katsaggelos, ''A compressed video enhancement algorithm,'' *Proc. IEEE ICIP'99*, pp. 454–458 (1999).
21. L. Boroczky and J. G. W. M. Janssen, ''Sharpness enhancement for MPEG encoded/transcoded video sources,'' *Proc. IEEE ICIP'01*, pp. 407–410 (2001).
22. Y. Yang and L. Boroczky, ''A new enhancement method for digital video applications,'' *IEEE Trans. Consum. Electron.* **48**(3), 435–442 (2002).
23. S. Mei-Yin and C. C. J. Kuo, ''Review of post processing techniques for compression artifact removal,'' *J. Vis. Commun. Image Rep.* **9**(1), 2–14 (1998).
24. H. C. Reeve and J. L. Lim, ''Reduction of blocking effect in image coding,'' *Opt. Eng.* **23**(1), 34–37 (1984).
25. C. M. Liu, V. N. Wang, and J. Y. Lim, ''A new postprocessing method for block-based DCT coding based on the convex-projection theory,'' *IEEE Trans. Consum. Electron.* **44**, 1054–1061 (1998).
26. T. S. Liu and L. W. Chang, ''An adaptive temporal-spatial filter for MPEG coded video signals,'' *Multidimensional Syst. Sign. Process.* **6**, 251–262 (1995).
27. T. Meier, K. N. Ngan, and G. Crebbin, ''Reduction of blocking effect in image and video coding,'' *IEEE Trans. Circuits Syst. Video Technol.* **9**(3), 490–500 (1999).
28. S. A. Martucci, ''A new approach for reducing blockiness in DCT image coders,'' *ICASSP 98*, pp. 2549–2552 (1998).
29. V. B. K. Konstantinides and G. Beretta, ''Image sharpening in the JPEG domain,'' *IEEE Trans. Image Process.* **8**(6), 874–878 (1999).
30. J. Tang and E. Peli, ''Image enhancement in JPEG domain for low-vision patients,'' *SID'02*, pp. 990–993 (2002).
31. J. Tang, E. Peli, and S. Acton, ''Image enhancement using a contrast measure in the compressed domain,'' *IEEE Signal Process. Lett.* **10**(10), 289–292 (2003).
32. ATSC RP A/54, ''Guide to the use of the ATSC digital television standard,'' see http://www.atsc.org/document_map/video.htm#A/54.
33. D. L. Gall, ''MPEG: A video compression standard for multimedia applications,'' *Commun. ACM* **34**(4), 47–58 (1991).
34. ''Introduction to MPEG 2 video compression,'' see http://www.bretl.com/mpeghtml/mpeg2vcl.HTM.
35. B. G. Haskell, A. Puri, and A. N. Netravali, *Digital Video: An Introduction to MPEG-2*, International Thomson Publishing (New York, New York, 1997).
36. C. Bowonkoon and K. R. Rao, ''Human visual weighted progressive image transmission,'' *IEEE Trans. Commun.* **38**(7), 1040–1044 (1990).
37. Restream Software, see http://shh.dvdboard.de.
38. MPEG elementary streams, see ftp://ftp.tek.com/tv/test/streams/Element/MPEG-Video/525.
39. Pixeltools Corp., see ftp://ftp.pixeltools.com.
40. ''Configuring TMPGEnc for high-quality, DVD-compliant MPEG-2,'' see http://dvd-hq.info/Compression.html.
41. MPEG Repair software, Pixeltools Corp., see http://www.pixeltools.com/MpegRepair.html.
42. E. Peli, E. Fine, and A. Labianca, ''The detection of moving features on a display: The interaction of direction of motion, orientation, and display rate,'' *SID'98*, pp. 1033–1036 (1998).

Jeonghoon Kim received his MS and PhD degrees in electronics from Yonsei University, Seoul, Korea, in 1989 and 2000, respectively. Since 1989 he has been with the Signal Processing Center at Samsung Electronics as a senior research engineer. From 2002 to 2003 he worked as a postdoctoral fellow at Schepens Eye Research Institute, Harvard Medical School, Boston. He is an assistant professor in the Department of Electronic Communication at the Shinheung College, Korea. His interests are vision rehabilitation, digital image processing, 3-D processing, and digital application specific integrated circuits (ASIC) design.

Avni Vora is currently a medical student at the University of Illinois-Chicago. She received her bachelor's degree in biomedical engineering at Duke University in 2002. Her research interests are in the use of digital technology in assisting visually impaired patients in both visual function and mobility.

Eli Peli received a BSEE, cum laude, and MSEE from the Technion-Israel Institute of Technology. He earned his doctorate from New England College of Optometry in Boston. He is a senior scientist and the Moakley Scholar in Aging Eye Research at The Schepens Eye Research Institute, and a professor of ophthalmology at Harvard Medical School. He also serves on the faculty of the New England College of Optometry (adjunct professor of optometry and visual sciences). Since 1983 he has been caring for visually impaired patients as the director of the Vision Rehabilitation Service at the New England Medical Center Hospitals in Boston. He is a Fellow of the American Academy of Optometry, a Fellow of the Optical Society of America, and a Fellow of the Society for Information Display. His principal research interests are image processing in relation to visual function and clinical psychophysics in low-vision rehabilitation, image understanding, and evaluation of display-vision interaction. He also maintains an interest in oculomotor control and binocular vision. He has published more than 100 scientific papers and has been awarded five U.S. patents. He edited a book entitled *Visual Models for Target Detection* with special emphasis on military applications, and most recently coauthored a book entitled *Driving with Confidence: A Practical Guide to Driving with Low Vision*.