CANCELLATION OF IMAGE CROSSTALK IN TIME-SEQUENTIAL DISPLAYS OF STEREOSCOPIC VIDEO

CANCELLATION OF IMAGE CROSSTALK IN TIME-SEQUENTIAL DISPLAYS OF STEREOSCOPIC VIDEO

Janusz Konrad, Bertrand Lacotte and Eric Dubois

Université du Québec
Institut national de la recherche scientifique
INRS-Télécommunications
16, place du Commerce, Verdun
Québec, Canada, H3E 1H6

Février 1997

Rapport technique de l'INRS-Télécommunications no. 97-01

Preliminary work on crosstalk cancellation was carried out by Bertrand Lacotte of ENST, Paris during his stay at INRS-Télécommunications between August and December 1995 and then described in a preliminary report [3]. The work was continued in 1996 by the other two authors.

This report is an exact replica of a paper submitted to the *IEEE Transactions on Image Processing* in February 1997.

The work presented here was supported by the Natural Sciences and Engineering Research Council of Canada and by the Fonds pour la formation de chercheurs et l'aide à la recherche (FCAR), Québec, Canada.

Summary

Stereoscopic visualization systems based on liquid crystal shutter (LCS) eyewear and cathode-ray tube (CRT) displays provide today the best overall quality of 3-D images and therefore have a dominant position in commercial as well as professional markets. Due to the CRT and LCS characteristics, however, such systems suffer from perceptual crosstalk ("shadows") at object boundaries that can reduce, and at times inhibit, the ability to perceive depth. In this paper, we propose a method to reduce such crosstalk. We present a simple model for intensity leak, we assess model parameters for a time-sequential LCS/CRT system and we propose a computationally-efficient algorithm to eliminate the crosstalk. Since the full crosstalk elimination implies an unacceptable image degradation (reduction of contrast), we study the trade-off between crosstalk elimination and image degradation. We describe experiments on synthetic and natural stereoscopic images and we discuss informal subjective viewing of processed images. Overall, the viewer response has been very positive; 3-D perception of many objects became either much easier or even effortless. Since the proposed algorithm can be easily implemented in real time (only linear scaling and table look-up are needed), we believe that it can be successfully used today in various stereoscopic applications suffering from image crosstalk.

Contents

List of Figures

1 Introduction

The recent increase of interest in stereoscopic and 3-D imaging can be traced to two sources. First, stereoscopic imaging has been successfully used for a number of years in such applications as medicine (3-D microscopy, ultrasound, training), science (3-D visualization) and remote guidance (remote manipulation using 3-D imaging feedback). Today, however, stereoscopic imaging is finding its way into homes *via* computer games and is expected to dramatically change TV, multimedia and electronic cinema in the near future. Secondly, the electronics industry is devoting a great deal of effort to the development of maximally realistic visual communications, high-definition TV (HDTV) activity being the prime example. The next logical step to increase the realism of visual communications is to include the 3-D (depth) information [6]. This is usually achieved by means of time-sequential display with active glasses or polarized projection with passive glasses [5]. Other technologies, such as autostereoscopic displays, have not been widely accepted yet, primarily due to the inferior quality that they have offered to date [7]. The recent developments in autostereoscopic non-lenticular displays [2] may, however, change this situation in the future.

In a time-sequential stereoscopic display, active LCS glasses are used for view separation; when the left image is presented on the screen, the left shutter is open whereas the right one remains closed [5]. When the right image arrives, the roles are reversed. To assure no flicker, each sequence (left and right) needs to be displayed at the original frame rate (50 or 60 Hz); the combined left-right sequence is then displayed at 100 or 120 Hz. A time-sequential presentation on a CRT monitor provides a good-quality 3-D perception, is relatively inexpensive for small installations (a 120 Hz monitor and a few pairs of LCS glasses) and easy to set up (no optical alignment needed, easily transportable, relatively immune to vibrations, shock, etc.). To reduce the cost of a time-sequential system, polarized passive glasses can be employed to view a CRT monitor via an LCS polarizing screen, but the quality is inferior to that of the direct viewing by LCS glasses. Alternative systems use polarized projection to separate views; images are synchronous in time but overlayed spatially. Although cost-effective for large installations (polarized glasses are far less expensive than the LCS ones), they require careful alignment of images and therefore are very sensitive to misplacement. While film-based projection systems are in daily use in IMAX-3D theatres, electronic projection setups are still experimental. Recently, LCS glasses have been used with a great success to view large-screen film in the newest IMAX-3D theatres. It seems, however, that it will be the electronic projection combined with active LCS glasses that will lead to cost-effective high-quality stereoscopic cinema.

Although in time-sequential systems the motion-parallax relationship is somewhat distorted (synchronously acquired left and right images are displayed 1/120-1/100 sec apart), the effect is not too serious. A more serious problem arises when LCS glasses are used to view CRT monitors; these are the least expensive and therefore the most popular setups, especially used in research, scientific visualization and at home (for

example "SimulEyes VR" from StereoGraphics Corp.). In such setups, the crosstalk-induced "shadows" are not only annoying but also reduce, or even inhibit, the viewer's ability to correctly perceive depth. In fact, we have observed that if a bright object is presented over a dark background close to a screen edge, the "shadows" seriously affect the viewer's ability to fuse the two cues into a meaningful 3-D object. This causes a substantial discomfort and may result in viewer bias against the LCS-based stereoscopic display technology.

There are two sources of the crosstalk described above:

1. *phosphor persistence*: by doubling the display frame rate with unchanged screen phosphors, the residual image intensity "leaking" into the subsequent image is significantly increased resulting in "shadows" if objects in consecutive images are not perfectly registered (motion, parallax),
2. *LCS characteristics*: liquid crystal shutters do not close completely; a fraction of light from the unintended image reaches each eye, e.g., when the right shutter is closed a fraction of the left stimulus reaches the right eye that is still reacting to the preceding right stimulus (temporal properties of the eye).

Although theoretically the first problem could be solved by employing new faster phosphors, it is difficult to find such phosphors without sacrificing certain screen characteristics (e.g., brightness) at standard refresh rates. Consequently, most of the "stereo-ready" monitors today use standard phosphors employed in ordinary 50/60/72 Hz monitors. See [1] for a more detailed discussion of phosphor persistence in time-sequential stereoscopic displays. As for the shutters, their impact is less severe [4] but not negligible. Although new-generation LCS glasses will offer a wider dynamic range, there are thousands of older models with higher leaks.

An interesting approach to the reduction of crosstalk between stereoscopic views has been proposed by Lipscomb and Wooten [4]. It is based on the principle of creating "anti-crosstalk"; images are pre-distorted in such a way that upon display shadows are largely suppressed. Lipscomb and Wooten have proposed a heuristic algorithm for "anti-crosstalk" generation that is relatively complicated and requires substantial computing power. Our approach is based on the same principle of creating "anti-crosstalk", however we study the crosstalk phenomenon by using an explicit crosstalk model. We carry out a psychovisual experiment to quantify the crosstalk as a function of both unintended and intended stimuli. Having found model parameters, we solve a system of non-linear equations and create a look-up table that is later used to generate the "anti-crosstalk". To assure an adjustable crosstalk compensation we perform either linear or non-linear mapping prior to table lookup; effectiveness of the compensation can be varied from a partial cancellation to a full elimination. While in the former case 3-D perception can be significantly improved with negligible image degradation, in the latter it is effortless at the cost of reduced contrast and increased brightness of the image. Overall, viewers' response has been very encouraging as they claimed an improved 3-D perception and reduced visual fatigue. The proposed algorithm is very fast as it only requires scaling and table look-up, and can be easily

implemented in real time by very modest hardware.

The paper is organized as follows. Section 2 presents the crosstalk model used, while Section 3 describes how parameters of that model were measured. In Section 4 an algorithm for crosstalk cancellation is developed and in Section 5 experimental results are discussed. Section 6 discusses issues related to the implementation and possible applications.

2 Crosstalk model

We assume a simple crosstalk model: in addition to the intended stimulus, a fraction of the unintended stimulus is perceived by each eye. Initial experiments, however, quickly proved that the fraction perceived also depends on the intended stimulus. For example, similar magnitude of "shadows" may be perceived in both of the following scenarios:

1. large unintended stimulus superposed over large intended stimulus,
2. small unintended stimulus superposed over small intended stimulus.

Thus, our model depends on both intended and unintended stimuli, and can be expressed as follows:

$$f(\mathbf{x}) + \varphi(f'(\mathbf{x}), f(\mathbf{x})) = \tilde{f}(\mathbf{x}), \tag{1}$$

where $\mathbf{x}$ denotes pixel coordinate, f is the intended image (stimulus) for a given eye, f' is the unintended image and $\tilde{f}$ is the image perceived by the eye. φ is a *crosstalk function* that quantifies the amount of crosstalk seen by an eye as a function of intended and unintended stimuli. Since our model applies only to a particular pixel and ignores the point spread function of the CRT and LCS, we will omit the dependence of f on $\mathbf{x}$ in the sequel.

For the left and right images f_l and f_r, the model can be written as follows:

$$\begin{aligned} f_l + \varphi(f_r, f_l) &= \tilde{f}_l, \\ f_r + \varphi(f_l, f_r) &= \tilde{f}_r, \end{aligned} \tag{2}$$

where f_l, f_r denote the original images (as seen on a regular monoscopic display) and $\tilde{f}_l, \tilde{f}_r$ are images with crosstalk (as perceived by a viewer on a stereoscopic display compliant with model (1)).

Since, in general, images contain color information, the model (2) needs to be applied in an adequate color space. Digital stereoscopic sequences are often provided in the *Y-U-V* (ITU-R 601) format, but it is not obvious how to evaluate the crosstalk function φ for chrominance components U and V. Since video monitors use additive color reproduction based on R, G and B primary colors, we apply model (2) in the *R-G-B* color space. After conversion of *Y-U-V* images to the *R-G-B* format, we apply crosstalk functions φ_R, φ_G, φ_B to each channel independently. The channel-independent model is appealing since in a typical monitor R, G and B phosphors

have different characteristics; for example, the green phosphor's persistence is longer than that of the red and blue phosphors.

3 Psychovisual characterization of the crosstalk

To evaluate the crosstalk functions φ_R, φ_G, φ_B we have carried out a psychovisual experiment for R, G, B components independently. The experiment calibrates our display system consisting of *CrystalEyes* LCS glasses from StereoGraphics Corp. (together with graphics display controller GDC-3) and *GDM-20E01T* multi-sync monitor from Sony Corp. A similar experiment would need to be performed for other setups. The experiment was performed as follows:

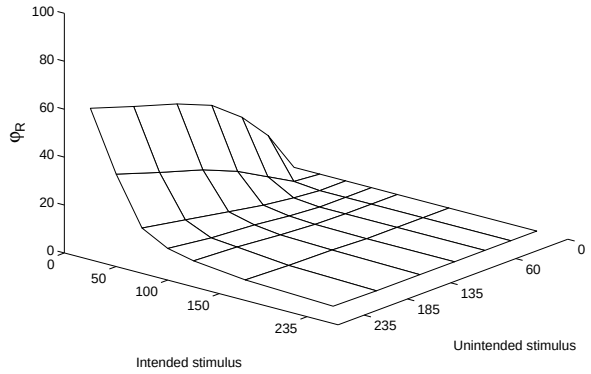
1. to minimize the impact of screen non-uniformity only the central part of the screen (20cm×10cm) was used; it was vertically divided into left and right halves,
2. the right eye was excluded from the experiment (the right shutter was permanently covered); we assumed that the results would be valid for the right eye as well,
3. only constant (spatially and temporally) primary-color stimuli (R, G and B) were used,
4. in the left half of the screen, a viewer-adjustable stimulus f_{adj} intended for the left eye was shown; no unintended stimulus was shown in that half,
5. in the right half of the screen, a stimulus f_l intended for the left eye was shown; at the same time a stimulus f_r unintended for the left eye (right-eye stimulus) was presented, thus inducing crosstalk,
6. the values of f_l and f_r were operator-selected in the $0-235$ range and the viewer was asked to match the two halves of the screen by adjusting f_{adj}.

Mathematically, the above description can be written as follows:

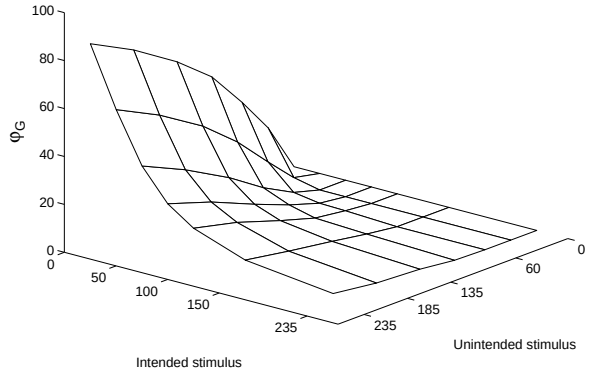
$$\begin{aligned} \text{left half:} \quad & f_{adj} + \varphi(0, f_{adj}) = f_{adj} + 0 = f_{adj} \\ \text{right half:} \quad & f_l + \varphi(f_r, f_l) = \tilde{f}_l \end{aligned}$$

where $\varphi(0, f_{adj})$=0 since no unintended stimulus was present in the left half of the screen. Once the viewer has matched the two parts of the screen, i.e., $\tilde{f}_l = f_{adj}$, we have $\varphi(f_r, f_l) = f_{adj} - f_l$. Since the number of matches to execute is equal to 3 (3 primary colors) times the product of the number of levels selected for f_l and f_r, we needed to judiciously select those levels in order to minimize viewer fatigue. After some initial experiments we realized that the surface φ exhibits larger curvature for smaller f_l (intended stimulus) and f_r (unintended stimulus). Thus, we selected a set of 7 nonuniform levels for f_l and another set of 6 nonuniform levels for f_r as follows:

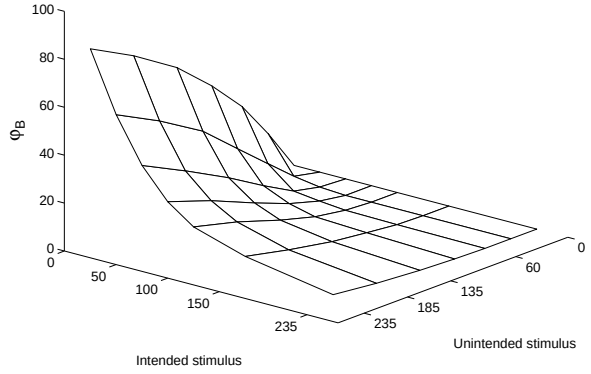
1. intended stimulus f_l: 0,25,50,75,100,150,235,
2. unintended stimulus f_r: 30,60,95,135,185,235.



(a) φ_R



(b) φ_G



(c) φ_B

Figure 1: R,G,B crosstalk functions psychovisually evaluated on a 7×6 grid (see text): (a) φ_R; (b) φ_G; and (c) φ_B. The bilinear facet representation shown is but one possible approximation of the crosstalk function between the measurement points.

Note that for f_r=0 no crosstalk is present and no evaluation is needed. Still, 126 matches had to be performed by each viewer. The initial experiment (green only) was performed by 5 viewers but since the results were very similar (often identical) we used only 3 viewers in the complete set of tests. During the experiments we noticed that the liquid crystal shutters do not have spatially-uniform transparency; intensity leakage is stronger at the LCS perimeter. This non-uniformity is very striking when examined with spatio-temporally constant stimuli, but is less obvious for complex images.

Crosstalk functions evaluated for each primary color are shown in Fig. 1; the results apply only to the central part of the screen viewed through the center of the shutters. The crosstalk functions for green and blue primaries are very similar (the amplitude of φ_G is slightly higher). The φ_R, however, attains lower amplitude and its surface is more curved than that of φ_G and φ_B. Note that for large values of the intended stimulus, the crosstalk is very small regardless of the unintended stimulus; any object over bright background induces little crosstalk. For small values of the intended stimulus, however, the crosstalk gets larger for larger values of the unintended stimulus; over a dark background, the brighter the object the stronger the crosstalk.

4 Algorithm for crosstalk elimination

Equations (2) describe a continuous mapping $\mathcal{T}: R^2 \to R^2$ that transforms (f_l, f_r) into $(\tilde{f}_l, \tilde{f}_r)$. Let the domain and range of $\mathcal{T}$ be $\mathcal{D}(\mathcal{T})$ and $\mathcal{R}(\mathcal{T})$, respectively.

Since the crosstalk φ is a continuous function that has been measured in a psychovisual experiment over a 7×6 grid, a model is needed to interpolate the value of the crosstalk for any unintended and intended stimuli. We have chosen the bilinear interpolation for its simplicity and for the small spatial support of its kernel, particularly important because of the small number of measurement points. We have also tested a least-squares approximation by a 2-D polynomial [3]. Although a 4-th order polynomial (15 coefficients) resulted in a maximum approximation error of less than 2, it was not monotonic in $\mathcal{D}(\mathcal{T})$; φ increased for high intended stimuli. At the same time, a lower-order polynomial resulted in a much higher approximation error.

To match the full dynamic range of 8-bit pixels we have bilinearly extrapolated each crosstalk function from Fig. 1 into the $[235, 255]$ range. With the above extrapolation $\mathcal{D}(\mathcal{T}) = [0, 255] \times [0, 255]$. To demonstrate the impact of the crosstalk on $(f_l, f_r) \in \mathcal{D}(\mathcal{T})$, Fig. 2 shows $\tilde{f}_l$ as a function of (f_l, f_r) computed from equation (2) for the extrapolated φ_B. Note the deviation from a plane for small values of the intended stimulus (f_l), especially for large values of the unintended stimulus (f_r); with no crosstalk, this should be a slanted plane crossing the unintended stimulus axis.

Another graphical representation of $\mathcal{T}$ that is easier to interpret is shown in Fig. 3. Fig. 3(a) shows a regular grid of values in the domain $\mathcal{D}(\mathcal{T})$, whereas Figs. 3(b-d) show the same grid after a mapping by $\mathcal{T}$ using the extrapolated φ_R, φ_G, φ_B. Note that $(0, 0)$ always maps onto $(0, 0)$, while other points undergo continuous but non-uniform

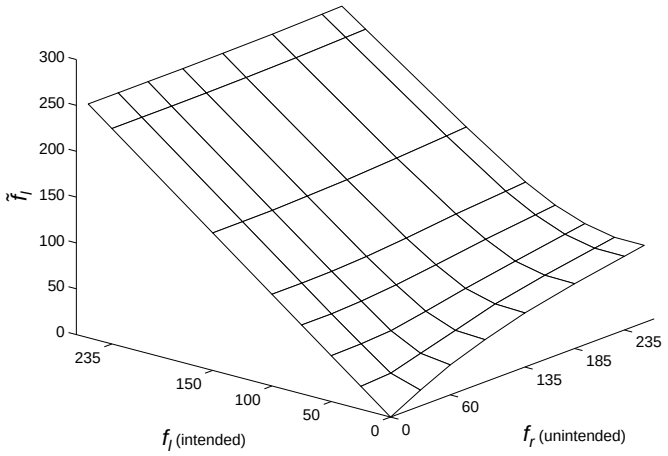


Figure 2: Example of mapping $\mathcal{T}$ for component B: $\tilde{f}_l = f_l + \varphi_B(f_r, f_l)$ for $(f_l, f_r) \in \mathcal{D}(\mathcal{T})$. Exactly the same mapping applies to $\tilde{f}_r$ but the roles of f_l and f_r are reversed.

warping. Also, note that the range $\mathcal{R}(\mathcal{T})$ does not cover the full $[0, 255] \times [0, 255]$ area. Clearly, the inverse mapping $\mathcal{T}^{-1}$ is not defined in $\Upsilon = \Upsilon_l \cup \Upsilon_r = [0, 255] \times [0, 255] \backslash \mathcal{R}(\mathcal{T})$ (the area between the dashed lines and the outermost curve of the transformed grid in Fig. 3(b-d)). Should the algorithm implementing $\mathcal{T}^{-1}$ be applied to points in Υ_l and Υ_r, negative values of f_l and f_r, respectively, would result, that cannot be displayed.

Knowing the original images (f_l, f_r) as well as the parameters of the crosstalk model φ, the goal is to find such images γ_l and γ_r that:

$$\begin{aligned} \gamma_l + \varphi(\gamma_r, \gamma_l) &= f_l, \\ \gamma_r + \varphi(\gamma_l, \gamma_r) &= f_r. \end{aligned} \tag{3}$$

Clearly, γ_l, γ_r are pre-processed versions of the original images such that after crosstalk superposition they become shadow-free. In other words, knowing the amounts of the crosstalk introduced by the visualization equipment, we want to remove these amounts from the original images. The effectiveness of crosstalk compensation will depend primarily on the accuracy of the proposed model as compared to the true leakage processes exhibited by screen phosphors and liquid crystal shutters (e.g., LCS spatial non-uniformity).

To compute (γ_l, γ_r), we have to know the inverse mapping $\mathcal{T}^{-1}$. Since the bilinear interpolation of φ is equivalent to a linear but shift-variant operator (FIR filter), it is not clear how to find a closed-form for the inverse mapping $\mathcal{T}^{-1}$. Note, however, that due to a typical 8-bit quantization of the original images (f_l, f_r), $\mathcal{T}^{-1}$ is only needed for 256 levels of each component's left and right pixels. In fact, due to the area Υ, where $\mathcal{T}^{-1}$ is undefined, the mapping $\mathcal{T}^{-1}$ is needed for many fewer pixel values. Thus, $\mathcal{T}^{-1}$ can be pre-computed and stored in a look-up table (LUT); a table for $256 \times 256 \times 3 \times 2$ intensities suffices. Although, in general, each "inverse" intensity is

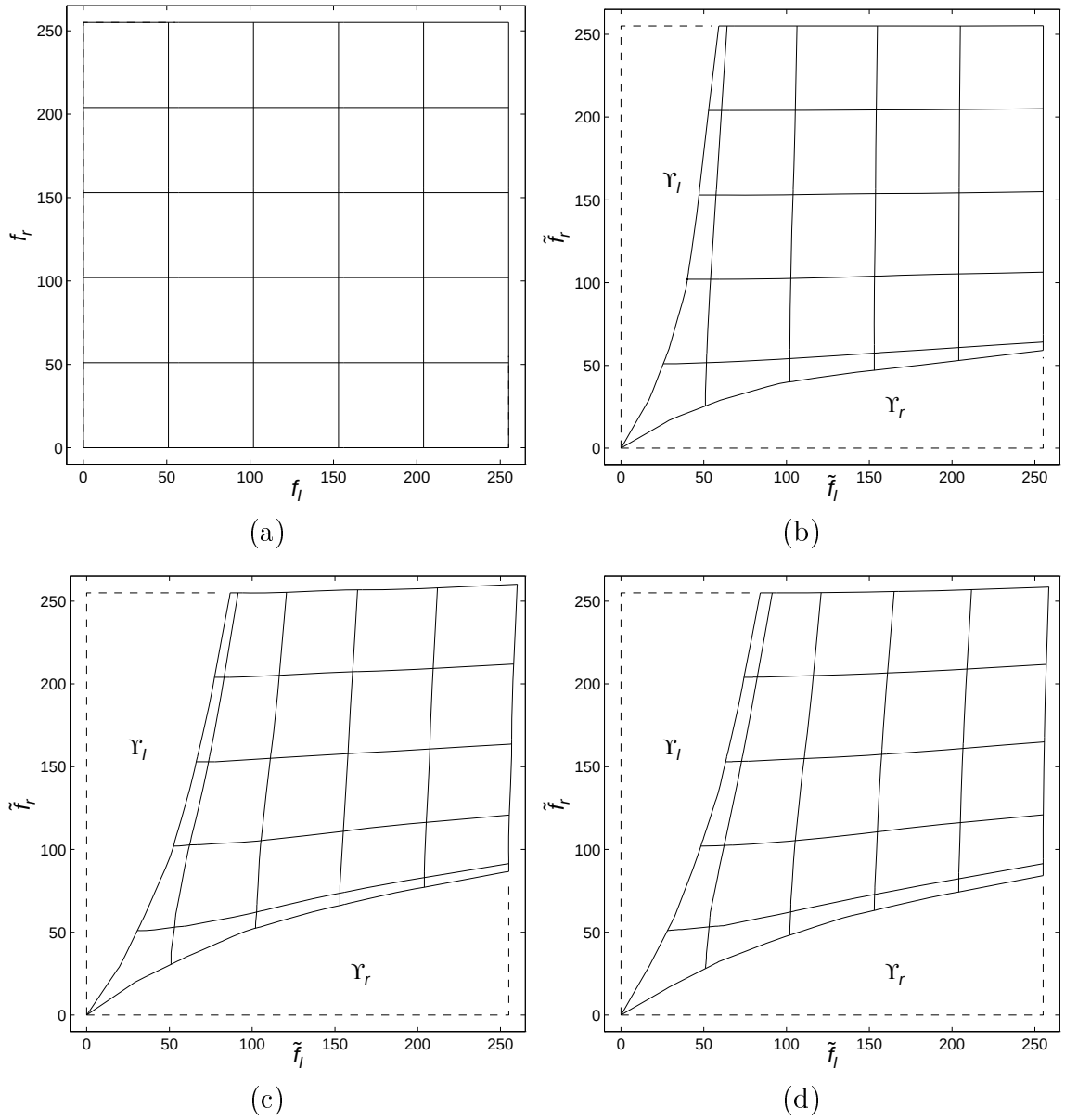


Figure 3: (a) Regular grid of intensities $(f_l, f_r) \in \mathcal{D}(\mathcal{T})$ with levels 0, 51, 102, 153, 204, 255, and the same grid after mapping by $\mathcal{T}$ (equation (2)) for (a) φ_R; (b) φ_G; (c) φ_B. See text for the meaning of Υ_l and Υ_r

a real number, we can round each number to an integer value in the $[0, 255]$ range since the human visual system cannot discern such small intensity differences. Therefore, less than 400kB of memory is needed to store the LUT, and if the area Υ is taken into account, the memory needed would be even smaller.

Since the crosstalk functions φ are monotonic and continuous (interpolation), we use the following iterative procedure to compute $\mathcal{T}^{-1}$:

$$\begin{aligned} \gamma_l^{(n+1)} &= f_l - \varphi(\gamma_r^{(n)}, \gamma_l^{(n)}), \\ \gamma_r^{(n+1)} &= f_r - \varphi(\gamma_l^{(n)}, \gamma_r^{(n)}), \end{aligned} \tag{4}$$

where the superscript n denotes the iteration number. The original pixel values are used to initiate γ, i.e., $\gamma_l^{(0)} = f_l$ and $\gamma_r^{(0)} = f_r$. We have tested the convergence of this algorithm by first applying the model (2) to integer values of R, G, B components in the range $[0, 255]$ and subsequently by using the resulting values as the input to (4). Since the rate of convergence of the algorithm depends on the gradient of φ, the slowest convergence was obtained for the R component (Fig. 1) for small values of the intended stimulus and large values of the unintended stimulus; φ_R is almost constant in the direction of the unintended stimulus. Nevertheless, a convergence to within less than 0.5 of each component value was attained after about 500 iterations. Since this computation needs to be carried out only once for a given set of crosstalk functions (the results are stored in a LUT), its complexity is not important.

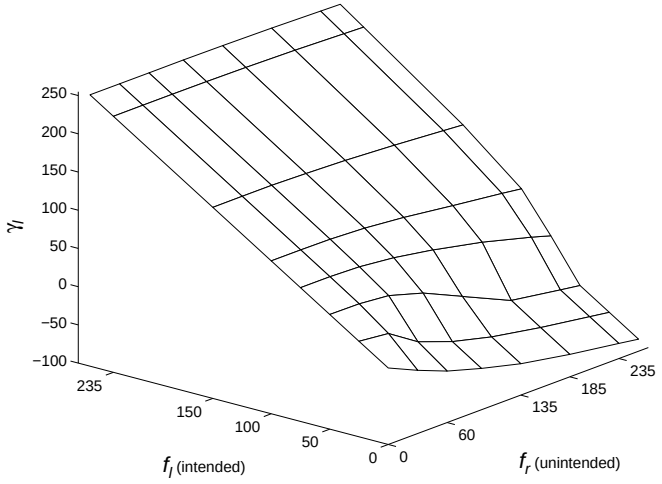
An example of the application of the above algorithm for φ_B is shown in Fig. 4; this is a complementary mapping to the one from Fig. 2. Note the compensating shape of the transformation for small values of the intended stimulus. Since most values of γ_l for large unintended stimulus and small intended stimulus are negative, $\mathcal{T}^{-1}$ is not defined there. The best that can be done there is to use the intensity non-negativity constraint and map the resulting negative values to zero. Fig. 4(b) presents γ_l computed under such a constraint and shown as intensity.

Note that an algorithm based on such a constraint is particularly simple. It requires two 256×256 matrices from Fig. 4(b) for each color component [1]; operations other than table look-up are not needed. As will be described later, the constraint on non-negativity of the intensity does not allow a full compensation of the crosstalk but maintains contrast and average brightness of the original image.

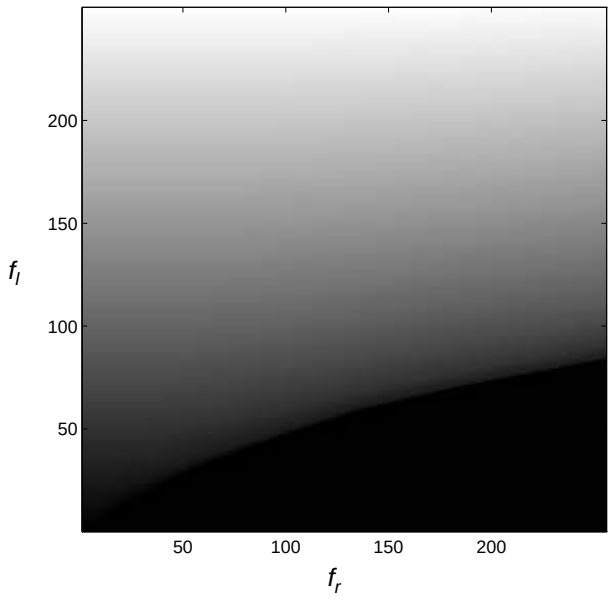
It is clear from Figs. 3(b-d) that only intensity pairs falling into the grid area $\mathcal{R}(\mathcal{T})$ can be fully compensated for crosstalk (as described by the model (2)). Thus, in order that the inverse mapping $\mathcal{T}^{-1}$ produce intensities in the $[0, 255] \times [0, 255]$ square, its domain must be $\mathcal{D}(\mathcal{T}^{-1}) = [0, 255] \times [0, 255] \backslash \Upsilon$. However, the original images f_l and f_r will certainly contain intensities outside the grid area, i.e., in Υ. To assure a full crosstalk cancellation, intensities in Υ must be mapped onto $\mathcal{D}(\mathcal{T}^{-1})$ before the compensation.

Let $\mathcal{T}(0, f_r) = (\psi(\tilde{f}_r), \tilde{f}_r)$ and $\mathcal{T}(f_l, 0) = (\tilde{f}_l, \psi(\tilde{f}_l))$. These are boundary curves of the transformed grid in Figs. 3(b-d) that neighbor Υ_l and Υ_r, respectively. Also,

[1] The 6 mapping matrices (Fig. 4(b)) can be made available to an interested reader.



(a)



(b)

Figure 4: Result of the application of algorithm (4) to the B component of left image, i.e., a solution γ_l to equation $f_l = \gamma_l + \varphi_B(\gamma_r, \gamma_l)$ for $(f_l, f_r) \in [0, 255] \times [0, 255]$: (a) surface plot; (b) representation as intensity. Since $\mathcal{T}^{-1}$ is defined in $\mathcal{R}(\mathcal{T})$, the surface plot represents $\mathcal{T}^{-1}$ for $\gamma_l \geq 0$ only while in the representation as intensity negative values have been set to zero. Exactly the same mapping applies to γ_r but the roles of f_l and f_r are reversed.

let $\mathcal{S}(\alpha)$ be a variable-size square in the $(\tilde{f}_l, \tilde{f}_r)$ plane defined by $[\alpha, 255] \times [\alpha, 255]$, where $0 \leq \alpha \leq \alpha_{max}$ and α_{max}=$\psi(255)$ for each primary color (Fig. 3). For α=α_{max} full crosstalk compensation is assured since the square $\mathcal{S}$ is included in the grid area, i.e., $\mathcal{S}(\alpha_{max}) \subset \mathcal{R}(\mathcal{T})$.

We have investigated pre-processing *via* linear and non-linear mappings defined as follows:

1. linear mapping $[0, 255] \times [0, 255] \rightarrow \mathcal{S}(\alpha)$:

$$f_l \leftarrow \alpha + f_l \frac{255 - \alpha}{255},$$
$$f_r \leftarrow \alpha + f_r \frac{255 - \alpha}{255},$$

2. non-linear mapping $[0, 255] \times [0, 255] \rightarrow \mathcal{S}(\alpha)$:

$$f_l \leftarrow \max(f_l, \alpha),$$
$$f_r \leftarrow \max(f_r, \alpha).$$

Clearly, both mappings reduce the dynamic range of the image by α; whereas the first one increases the average intensity, the second one eliminates dark-area details.

Note that for a fixed α the above mappings can be easily incorporated into LUT discussed in Section 4 and therefore will not add to the complexity of a real-time implementation. A more flexible solution would be to permit a user-adjustable α. This would be possible either by performing the mapping in real time before table look-up (increased complexity) or by precomputing the LUT off-line after each change of α (no immediate feedback upon changing α).

5 Experimental results

We have applied the algorithms described above to the CCETT stereoscopic video sequences "piano", "tunnel", "train", "manege" [2], to the NHK sequences "flower" and "streetorgan" as well as to our own computer-generated sequence of two moving spheres. Since the crosstalk model and the compensation algorithm apply to primary-color channels, the original sequences in *Y-U-V* (4:2:2) format were converted to the *R-G-B* (4:4:4) format, processed to eliminate crosstalk and then converted back. A simple linear interpolation was used for chrominance up-conversion, while due to narrow chrominance bandwidth, subsampling with no prefiltering was applied during the down-conversion. The algorithm is very fast because the processing is limited to a table look-up; image access from a disk and format conversion take much more time than the crosstalk compensation itself.

Although the crosstalk functions φ have been calibrated for Sony's *GDM-20E01T* monitor, the tests described below have been carried out on both the Sony monitor

[2] See Acknowledgments section.

and a NEC *XP 17* stereo-ready monitor. We have detected no subjective difference in crosstalk compensation between the two monitors. In fact, contrast and brightness adjustments had more impact on the compensation than the choice of monitor. The best crosstalk compensation for the Sony monitor was achieved with the same calibration as used during the measurements of crosstalk functions. During tests of the NEC monitor, contrast and brightness were slightly readjusted.

First, we tested the crosstalk compensation algorithm with the linear mapping. Although the crosstalk was completely eliminated for α=α_{max} (α_{max}=59, 87, 85 for R, G, B, respectively), processed images were much brighter than the originals and the contrast was markedly reduced. The images, however, could be effortlessly fused and observers have claimed significantly improved 3-D perception due to the elimination of "shadows".

Then, we tested the non-linear mapping (saturation) also with α=α_{max} in each channel. In this case, although the dynamic range was reduced as well, only detail in dark image areas was lost. At the same time dark areas were "whitened" since intensity values below α_{max} were "pushed" to α_{max}; dark areas looked as if they were watched through a glass window with a light reflection. Although "shadows" vanished, the solution was unacceptable as well.

Consequently, we tested smaller values of α that would result in a residual crosstalk but would substantially reduce the above effects. Since for $\alpha < \alpha_{max}$ the intersection $\mathcal{S}(\alpha) \cap \Upsilon$ is not empty, intensities in this intersection are mapped to 0 rather than to negative values (intensity non-negativity constraint). We tested α between 10 and 60 with an increment of 10; the same α was used in each color channel. In informal subjective viewing we have found that the non-linear mapping with α=20-30 assured a substantial reduction of the crosstalk for all tested images while introducing only a slight dark-area detail loss and marginal "whitening". Consequently, 3-D perception was significantly enhanced. For example, "shadows" around the following objects have been significantly suppressed:

1. flower leaves in "piano",
2. red-and-white semaphore and post in "train",
3. red-and-white post, ladder and white helmets in "tunnel",
4. bus and bus stop in "manege",
5. girl's hat in "streetorgan".

The improvement in 3-D perception was particularly striking for the red-and-white semaphore in "train" and the red-and-white post in "tunnel" both near picture edge.

We have found that for α=20-30 linear mapping gives even better crosstalk suppression. Although the contrast was slightly reduced and the average image intensity increased noticeably, the crosstalk was almost fully compensated unlike in the case of the non-linear mapping (for the same α). The 3-D perception improved dramatically and observers have been very positive about the subjective value of the improvement. In fact, in all tested images only objects causing extreme contrast, such as the white helmet in the tunnel next to the locomotive in "tunnel" and the white features on

the black door in "streetorgan", require a much higher α (about 60) to be fully compensated. The penalty paid is the increased image brightness due to scaling – an acceptable solution in quality-oriented applications.

Finally, we have tested the crosstalk compensation algorithm with non-negativity constraint only (no pre-processing to map Υ). With this approach only partial crosstalk compensation was achieved; the 3-D perception was significantly improved, however it was inferior to linear and non-linear mappings with α=20-30. The advantage of this approach is that image fidelity is maintained (only slight loss of dark-area detail), while the crosstalk is largely suppressed. The approach may be of interest in applications were image fidelity is of primary concern.

To visually demonstrate how the compensation algorithm affects images, Fig. 5 shows a small luminance window from field #0 of the sequence "piano". In the original image silver-white leaves are presented over a blue background. In the bottom row are shown the same windows after the non-linear mapping with α=30. Note the dark imprints in the background of either window corresponding to bright leaves in the other window. It is exactly in these areas that the LCS leakage and phosphor persistence will get superposed onto the original image to provide a crosstalk-free perception.

There is an additional benefit of the above compensation. The overall brightness of each crosstalk-compensated image is reduced compared to the corresponding original image if viewed monoscopically; the same compensation principle resulting from equation (3) applies everywhere in the image. Therefore, compensated images viewed stereoscopically have a brightness corresponding to the monoscopic image, whereas uncompensated images are slightly brighter in the stereo mode due to the cross-view leak. This effect, however, is not serious enough to call for a compensation by itself.

6 Implementation and applications

We see possible applications for the proposed algorithm in stereoscopic visualization systems using a digital image format. An immediate application would be in computer-based stereo setups with substantial CPU power, e.g., powerful PCs or workstations; images would be pre-processed in software before shipping to a graphics board. This solution may require suitable conversions if the available data is not in the *R-G-B* format. Although for video sequences the needed CPU power would be substantial, still stereo images could be pre-computed and only then displayed. Also, user-adjustable α would requires extra CPU cycles.

Alternatively, the data can be processed in hardware on a graphics board after conversion to the *R-G-B* format but before D/A conversion; this is the only option for setups void of CPU power. No format conversion would be needed, but a 400kB memory and a simple scaling or "max" operator would have to be included on the board. Although the least expensive PC graphics boards do not satisfy these requirements, more advanced boards are very close to meeting them. With such boards, a

real-time adjustment of α could become possible as well.

It is important to note that the proposed crosstalk compensation algorithm can and should be adapted to a particular application by selecting a suitable type of mapping and its parameter α. Both should be user-selectable in real-time for a maximum visual comfort; for example, a reference sequence could be used to perform adaptation to both eyewear/screen setup and user's tolerance of crosstalk.

Based on our experience we believe that if fidelity to the original data is critical, no mapping should be applied (only non-negativity constraint) thus resulting in a partial crosstalk elimination. If a slight departure from the original (dark areas only) can be tolerated, then non-linear mapping with α=20-30 should be used; we believe that the incurred dark detail loss should not be objectionable to most viewers. However, if the benefits of effortless 3-D fusion and undisturbed perception of depth outweigh image contrast, we suggest to use the linear mapping with α=20-30; higher values should be used only in extreme cases. We believe that for 3-D TV, 3-D games or other quality-oriented stereoscopic applications the first two scenarios apply; no or almost unnoticeable change in contrast/brightness is introduced. However, for such applications as 3-D visualization, microscopy/ultrasound or remote guidance, we believe that the latter solution incurring brighter images is appropriate since the correct perception of depth is at least as important there as the faithfulness of color or intensity.

7 Summary and conclusions

We presented a unique approach to the enhancement of 3-D perception in time-sequential displays of stereoscopic images using LCS glasses. We proposed a simple crosstalk model, we evaluated its parameters in a psychovisual experiment and we proposed a computationally-efficient algorithm for crosstalk elimination. We demonstrated that a simple version of the algorithm using only the non-negativity constraint assures negligible distortion but only partial crosstalk compensation. By introducing a pre-processing stage in the form of a non-linear or linear mapping, we improved the crosstalk cancellation capacity at the cost of reduced contrast and increased image brightness. Since both mappings are controlled by a single parameter, a continuous adjustment between partial compensation/negligible degradation and full compensation/significant degradation can be performed by the user. The adjustment would depend on user preference in terms of the maximum visual comfort and on stereoscopic application used. We believe that the presented algorithm can be implemented today on an advanced graphics board, and therefore can be incorporated into various stereo applications such as 3-D microscopy, remote guidance or 3-D computer games.

Acknowledgments

We would like to thank Dr. Bruno Choquet of the CCETT, Rennes, France and the RACE DISTIMA project of the European Community for providing us with the stereoscopic sequences used in this work.

References

[1] P. Bos, "Time sequential stereoscopic displays: The contribution of phosphor persistence to the "ghost" image intensity," in *Proc. ITEC'91 Annual Conf., Three-Dimensional Image Tech.,* (H. Kusaka, ed.), pp. 603–606, July 1991.

[2] D. Ezra, G. Woodgate, B. Omar, N. Holliman, J. Harrold, and L. Shapiro, "New autostereoscopic display system," in *Proc. SPIE (Stereoscopic Displays and Virtual Reality Systems II)*, vol. 2409, pp. 31–40, Feb. 1995.

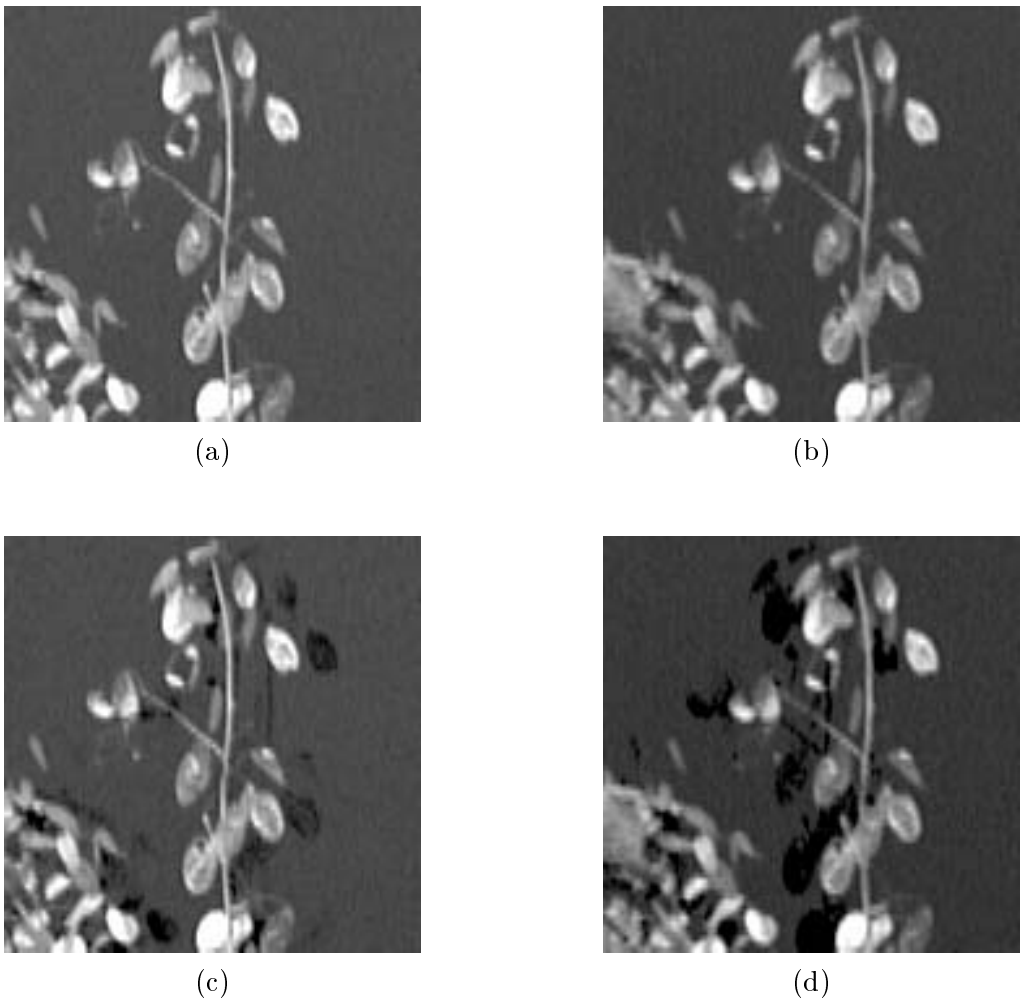
[3] B. Lacotte, "Elimination of keystone and crosstalk effects in stereoscopic video," Tech. Rep. 95-31, INRS-Télécommunications, Dec. 1995.

[4] J. Lipscomb and W. Wooten, "Reducing crosstalk between stereoscopic views," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 2177, Feb. 1994.

[5] L. Lipton, "Compatibility issues and selection devices for stereoscopic television," *Signal Process., Image Commun.*, vol. 4, pp. 15–20, 1991.

[6] L. Lipton, "True stereoscopic television: 3D-TV is feasible and striking," *Advanced Imaging*, pp. 28–30, July 1994.

[7] I. Sexton, T. Bardsley, and A. Bhoopal, "Errors in depth," in *Int. Workshop on Stereoscopic and 3D Imaging*, pp. 235–242, Sept. 1995.



(a) (b)

(c) (d)

Figure 5: Luminance of the original (a) left and (b) right fields #0 from "piano" sequence, and (c-d) the same images compensated for crosstalk using non-linear mapping with α=30. Since the sequence is 2:1 interlaced, vertical interpolation by 2 was applied to maintain correct aspect ratio.