

Cancellation of image crosstalk in time-sequential displays of stereoscopic video

Janusz Konrad, *Senior Member, IEEE*, Bertrand Lacotte and Eric Dubois, *Fellow, IEEE*

Abstract—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 contrast. 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. This is particularly true for PC-based 3-D viewing where the algorithm can be executed by the CPU or by an advanced graphics board.

Index Terms—Stereoscopic displays, stereoscopic and 3-D imaging, 3-D TV

I. 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) services being the prime example. The next logical step to increase the realism of visual communications is to include the 3-D (depth) information [7]. This is usually achieved by means of time-sequential display with

J. Konrad is with INRS-Télécommunications, 16 Place du Commerce, Verdun, QC, H3E 1H6, Canada (konrad@inrs-telecom.quebec.ca). E. Dubois is with the School of Information Technology and Engineering, University of Ottawa, Ottawa, ON, K1N 6N5, Canada (edubois@uottawa.ca). B. Lacotte worked on this project during a five-month training period at INRS-Télécommunications in 1995 on leave from Ecole Nationale Supérieure de Télécommunication, 46 rue Barrault, 75634 Paris Cedex 13, France.

This work was supported by the Natural Sciences and Engineering Research Council of Canada under research grant OGP121619 and by the Fonds pour la formation de chercheurs et l’aide à la recherche (FCAR), Québec, Canada under research grant 96-ER-1577.

active glasses or polarized projection with passive glasses [6]. Other technologies, such as autostereoscopic displays, have not been widely accepted yet, primarily due to the inferior quality that they have offered to date [8]. The recent developments in autostereoscopic non-lenticular displays [2] may, however, change this situation in the future.

The success of 3-D TV and 3-D multimedia will certainly depend on the ability to reduce the cost of a stereoscopic system and on the acceptance of the technology by viewers. While stereoscopic camera and display technology are still experimental and therefore expensive, stereoscopic transmission can be achieved today at no extra cost due to the Multi-View Profile in MPEG-2. As for the acceptance, concerns that 20-30% of the population suffers from stereo-blindness seem to have been greatly exaggerated. As shown recently, only about 4% of young viewers are stereo-anomalous if sufficiently long stimuli (over 1 second) are presented [10].

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 [6]. 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.). When the number of viewers grows, however, the cost of LCS glasses starts to dominate. In such a scenario, the cost-effective solution may be to employ polarized passive glasses and CRT monitor equipped with a polarizing screen (liquid crystal modulator). Alternative systems use polarized projection to separate views; images are synchronous in time but overlaid spatially with orthogonal polarizations (horizontal/vertical or clockwise/counter-clockwise). Although cost-effective for large installations (polarized glasses are far less expensive than the LCS ones), such systems are susceptible to image misalignment since, unlike in time-sequential displays, there is no inherent alignment between the two imaging channels. While film-based stereoscopic projection systems with either polarized or LCS eyewear are in daily use in Imax-3D[®] theatres, electronic projection setups are still experimental. 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 re-

lationship 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 and scientific visualization as well as at home (for example “SimulEyes VR” from StereoGraphics Corp., San Rafael, CA or “VR Visualiser” from VRex Inc., Elmsford, NY). 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 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 three primary sources of the crosstalk in time-sequential LCS/CRT-based systems:

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 (e.g., due to parallax). There are two components of phosphor persistence. The *fast-decay component* is responsible for increased crosstalk at the bottom of the screen; after the glasses switch during the vertical blanking, an afterglow of the recently-painted area (previous image) remains visible. The *slow-decay component*, although less pronounced, induces afterglow across the whole image; once a shutter is open, a residual of the previous (unintended) image remains visible throughout the current (intended) image. A detailed treatment of phosphor persistence can be found in [1].
2. *LCS leakage*: 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 (lowpass filtering in time of the human eye). Moreover, the spatial characteristics of the shutters are not uniform; usually, more crosstalk is visible at shutter perimeter.
3. *LCS shutter timing errors*: The optimal time to switch the shutters is during the vertical blanking, just before the raster starts drawing on the CRT (top of the screen). Any timing error, i.e., switching too early or too late, will induce additional crosstalk. For example, if switching is delayed, the top of new image is painted while the glasses finish to switch; increased crosstalk at the top of the screen results. This happens with most of the glasses since their switching time is longer than the vertical blanking interval.

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. Encouraging ex-

perimental results for P-43, a rare-earth green phosphor, replacing the standard P-22 phosphor have been presented in [1], but no trade-offs between crosstalk reduction and screen characteristics have been discussed. As for the shutters, their impact is less severe [5] but spatial non-uniformity of transparency/opaque characteristics can cause disturbing crosstalk at shutter perimeter. Although new-generation LCS glasses will offer a better dynamic range and thus reduce the impact of fast-decay phosphor persistence, widely-used older glasses will continue to suffer from higher levels of light leakage. The new glasses will not, however, reduce the crosstalk due to slow-decay persistence; this can only be achieved by new phosphors. Finally, the shutter timing errors may cause strong crosstalk at the top or bottom of the screen, but only over a few lines; using a slight overscan or reducing image size are but two simple, although not elegant, solutions.

An effective technique for the reduction of crosstalk induced by fast-decay phosphor persistence has been described in [1]. Since such crosstalk is visible at the bottom of the screen only, a staggered seamless switching has been proposed; shutter opening is executed in horizontal bands rather than instantaneously. At the beginning of the scan only a shutter band at the top of the screen is open; the afterglow at the bottom of the screen is not visible. The approach applies to shutters/polarizers with a fixed position with respect to the screen and therefore cannot be used in viewer-worn glasses. It has been used successfully in polarizing screens (e.g., 5-segment panel from NuVision Technologies Inc., Beaverton, OR and “ZScreen” from StereoGraphics Corp.).

A digital approach to the reduction of crosstalk (fast-decay and slow-decay) between stereoscopic views has been proposed by Lipscomb and Wooten [5]. It is based on the principle of creating “anti-crosstalk”; images are pre-distorted in such a way that upon display ghosting is largely suppressed. Lipscomb and Wooten have proposed a heuristic algorithm for “anti-crosstalk” generation that is relatively complicated and requires substantial computing power. More recently, a patent has been awarded to Street [9] for stereo image enhancement based on the same principle. In this patent, a very simple (two-parameter) linear crosstalk model has been proposed. The original images are pre-conditioned (addition of “anti-crosstalk”) by inverting a 2×2 matrix of the crosstalk model. However, no indication is given as to how model’s parameters should be measured.

Our approach, although developed independently [4] of the above work, is also based on the “anti-crosstalk” principle. However, unlike in the Lipscomb and Wooten’s case, we use an explicit crosstalk model. Moreover, unlike in the Street’s patent our crosstalk model is non-linear and accounts for the dependence of crosstalk on both intended and unintended stimuli. To quantify this complex model as a function of both stimuli, we carry out psychovisual experiments. Having found model parameters, we solve a system of non-linear equations and create a look-up table allowing 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 partial to full cancellation. While for the partial cancellation 3-D perception can be significantly improved with negligible image degradation, for

full cancellation the perception is effortless at the cost of reduced contrast and increased brightness of the image. Overall, viewers' response has been very encouraging with claims of 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. In fact, it can be implemented today directly on advanced graphics boards.

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

II. 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 ghosting may be perceived in both of the following scenarios:

1. high-brightness unintended stimulus superposed over high-brightness intended stimulus,
2. low-brightness unintended stimulus superposed over low-brightness intended stimulus.

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

$$f'(\mathbf{x}) = f(\mathbf{x}) + \phi(g(\mathbf{x}), f(\mathbf{x})), \quad (1)$$

where \mathbf{x} denotes pixel coordinate, f' is the image perceived by the eye, f is the intended (displayed) image for a given eye and g is the unintended image (source of crosstalk). ϕ is a *crosstalk function* that quantifies the amount of crosstalk seen by an eye as a function of unintended and intended 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 (f_l) and right (f_r) images, the joint model can be written as follows:

$$\begin{aligned} f'_l &= f_l + \phi(f_r, f_l), \\ f'_r &= f_r + \phi(f_l, f_r), \end{aligned} \quad (2)$$

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

The model (2) is scalar and is applicable to black-and-white images. To extend the model to color images, three crosstalk functions must be applied in a suitable color space. Since video monitors use additive color reproduction based on R, G and B primary colors, we apply the model in the $R-G-B$ color space; each of the three crosstalk functions ϕ_R, ϕ_G, ϕ_B characterizes one channel. The three-channel 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.

As we mentioned before, our approach is based on the creation of anti-crosstalk, and is similar to the method proposed by Lipscomb and Wooten [5]. There are fundamental differences between the two models, however. Whereas we evaluate the crosstalk function ϕ in a psychovisual experiment (Section III), the approach taken in [5] to assess the correction needed is heuristic, especially regarding the special treatment of thin vertical lines and left edges of bright objects. On the other hand, Lipscomb and Wooten allow for a vertically-variable anti-crosstalk to compensate for shutter timing errors (top of the screen) and fast-decay phosphor persistence (bottom of the screen). Once quantified, this variability could be easily incorporated into our algorithm; in our stereo setup there was no extra crosstalk at the top and bottom of the screen due to black horizontal bands above and below the image (reduced vertical size).

III. 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 calibrated our display system consisting of *CrystalEyes* LCS glasses from StereoGraphics Corp. (including GDC-3 graphics display controller) and *GDM-20E01T* multi-sync monitor from Sony Corp. Although theoretically a similar experiment would need to be performed for other setups, an application of psychovisual measurements for the Sony monitor to crosstalk cancellation on a NEC monitor worked very well (Section V).

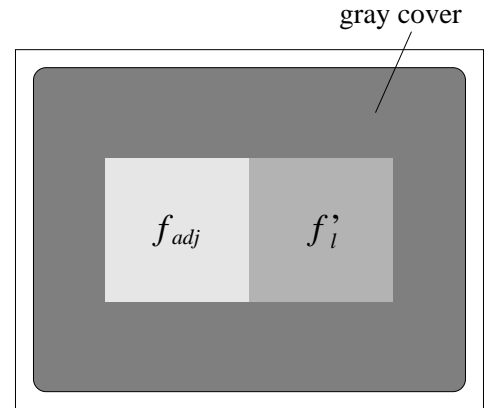


Fig. 1. Experimental setup for the psychovisual characterization of the crosstalk function for the left eye. f_{adj} is a viewer-adjustable stimulus intended for the left eye whereas f'_l is a combination of the stimulus f_l intended for the left eye and of f_r that is unintended for the left eye (right-eye stimulus). See text for details.

The experiment was performed as follows (Fig. 1):

1. to minimize the impact of screen and shutter non-uniformity on measurements only the central part of the screen was visible (20cm horizontally by 10cm vertically) and viewers were asked to view this area only through the center of the shutters; the visible area was vertically divided into adjacent (no separation) left and right halves (Fig. 1),

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 spatially- and temporally-constant 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 ($f_r=0$),
5. in the right half of the screen, both a stimulus f_l intended for the left eye and a stimulus f_r unintended for the left eye (right-eye stimulus) were presented, thus inducing crosstalk,
6. the values of f_l and f_r were operator-selected and the viewer was asked to match the two halves of the screen by adjusting f_{adj} .

Clearly, the experimental results should be valid only for the central part of the screen viewed through the center of the shutters. We have disregarded screen non-uniformity since in LCS-based systems it plays a secondary role to shutter non-uniformity that cannot be easily corrected since glasses do not have fixed relative position with respect to the screen. Perhaps, the only possibility is an improvement of shutter uniformity during manufacturing. Although the proposed approach to crosstalk cancellation is optimized for the center of the shutters, in general it is through this center that the viewer will fixate. Since for the viewer the fixation point belongs to the most important area of the image (at a given time instant), optimal crosstalk suppression should dominate. This is based on the assumption that to fixate the viewer rotates his head, perhaps an invalid assumption for small screens.

Mathematically, the description of the experimental setup can be written as follows:

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

where $\phi(0, f_{adj})=0$ 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., $f_l' = f_{adj}$, we have $\phi(f_r, f_l) = f_{adj} - f_l$. 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.

Note that for $f_r=0$ no crosstalk is present and no evaluation was needed. Since the crosstalk reaches saturation for high-brightness intended or unintended stimuli, to reduce the number of measurements we stopped at level 235; we bilinearly extrapolated the crosstalk value for higher levels. Still, 126 matches had to be performed by each viewer. The initial experiment (green only) was performed by 5 viewers but since the results

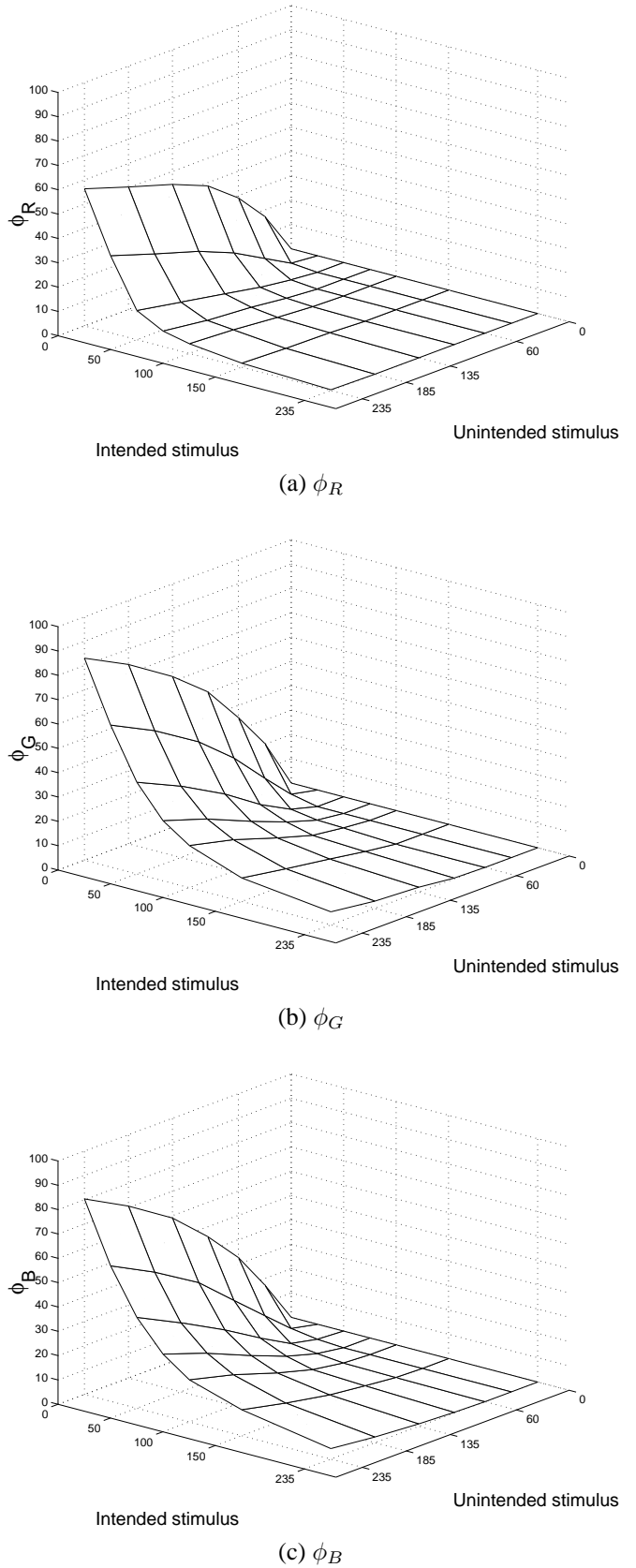


Fig. 2. R, G, B crosstalk functions psychovisually evaluated on a 7×6 grid (see text): (a) ϕ_R ; (b) ϕ_G ; and (c) ϕ_B . The bilinear facet representation shown is but one possible approximation of the crosstalk function between the measurement points.

were very similar (often identical) we used only 3 viewers in the complete set of tests. During the experiments we confirmed that the liquid crystal shutters do not have spatially-uniform transparency; intensity leakage was stronger at the LCS perimeter (clover leaf pattern). This non-uniformity was very striking when examined with spatio-temporally constant stimuli, but less obvious for complex images.

Crosstalk functions for each primary color computed as an average of the three measurements (3 subjects) are shown in Fig. 2; 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 crosstalk function, however, attains lower amplitude and its surface is more curved than that of ϕ_G and ϕ_B . Similar observation regarding smaller crosstalk in the red component has been made elsewhere [5]. 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.

Table I lists numerical values of the crosstalk functions; values shown are integer-rounded averages of measurements for 3 subjects and include extrapolated values at level 255. Non-rounded values as well as other information can be found on our web site [3]. The same table shows average, standard deviation and frequency of occurrence for the maximum spread (difference between maximum and minimum) among the measured values for the three viewers for each of the 42 data points and for each component (R , G and B).

IV. ALGORITHM FOR CROSSTALK ELIMINATION

Equations (2) describe a continuous mapping $T: R^2 \rightarrow R^2$ that transforms (f_l, f_r) into (f'_l, f'_r) . Let the domain and range of T be $\mathcal{D}(T)$ and $\mathcal{R}(T)$, 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 [4]. Although a 4-th order polynomial (15 coefficients) resulted in a maximum approximation error of less than 2, it was not monotonic over $\mathcal{D}(T)$; ϕ increased for high-brightness 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 (crosstalk must be known for all pixel values between 0 and 255), we have bilinearly extrapolated each crosstalk function from Fig. 2 into the $[235, 255]$ range. With the above extrapolation $\mathcal{D}(T) = [0, 255] \times [0, 255]$. To demonstrate the impact of the crosstalk on $(f_l, f_r) \in \mathcal{D}(T)$, Fig. 3 shows f'_l as a function of (f_l, f_r) computed from equation (2) for the extrapolated function 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.

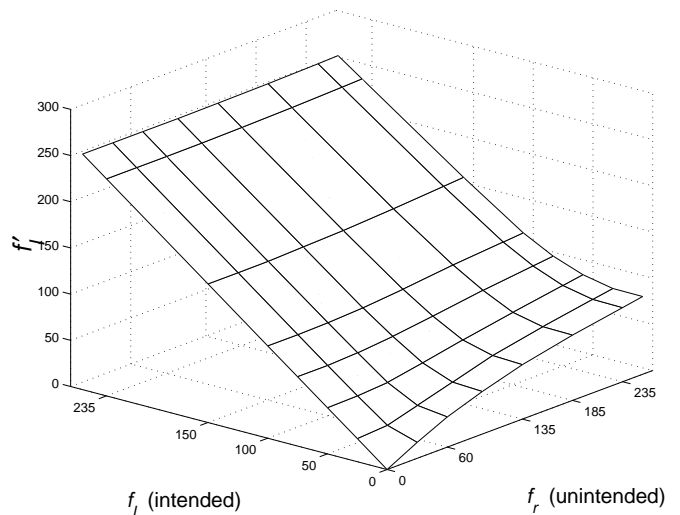


Fig. 3. Example of mapping T for component B : $f'_l = f_l + \phi_B(f_r, f_l)$ for $(f_l, f_r) \in \mathcal{D}(T)$. Exactly the same mapping applies to f'_r , but the roles of f_l and f_r are reversed.

Another graphical representation of T that is easier to interpret is shown in Fig. 4. Fig. 4(a) shows a regular grid of values in the domain $\mathcal{D}(T)$, whereas Figs. 4(b-d) show the same grid after a mapping by T using the extrapolated functions ϕ_R , ϕ_G , ϕ_B . Note that $(0, 0)$ always maps onto $(0, 0)$, while other points undergo continuous but non-uniform warping. Also, note that the range $\mathcal{R}(T)$ does not cover the full $[0, 255] \times [0, 255]$ area. Clearly, the inverse mapping T^{-1} is not defined in $\Upsilon = \Upsilon_l \cup \Upsilon_r = [0, 255] \times [0, 255] \setminus \mathcal{R}(T)$ (the area outside the warped region in Fig. 4(b-d)). Should the algorithm implementing 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:

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

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 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 T^{-1} . Note, however, that due to a typical 8-bit quantization of the original images

TABLE I

NUMERICAL VALUES OF CROSSTALK FUNCTIONS PSYCHOVISUALLY EVALUATED ON A 7×6 GRID: (A) ϕ_R ; (B) ϕ_G ; (C) ϕ_B , AND (D) STATISTICS OF THE MAXIMUM SPREAD AMONG THE MEASURED VALUES (SEE TEXT FOR DETAILS). THE VALUES SHOWN IN (A-C) ARE INTEGER-ROUNDED AVERAGES OF MEASUREMENTS FOR 3 SUBJECTS AND INCLUDE BILINEARLY EXTRAPOLATED VALUES AT LEVEL 255 THUS RESULTING IN AN 8×7 GRID.

(a)		(b)													
Intended	Unintended stimulus							Intended	Unintended stimulus						
	30	60	95	135	185	235	255		30	60	95	135	185	235	255
0	17	29	39	45	50	56	59	0	20	35	50	62	74	83	87
25	1	7	14	20	26	32	34	25	2	13	26	38	49	58	62
50	0	1	3	5	9	12	13	50	0	3	10	19	29	38	41
75	0	0	1	3	4	6	7	75	0	1	5	11	19	25	27
100	0	0	0	1	3	4	4	100	0	1	3	7	13	17	19
150	0	0	0	1	1	2	2	150	0	0	2	4	7	10	11
235	0	0	0	0	0	0	0	235	0	0	0	2	3	5	6
255	0	0	0	0	0	0	0	255	0	0	0	2	2	4	5

(c)		(d)													
Intended	Unintended stimulus							Average	Standard deviation	Freq. of occur. (%)					
	30	60	95	135	185	235	255			0	1	2	3	4	
0	17	33	46	59	71	80	84	<i>R</i>	1.05	1.23	43	31	12	7	7
25	2	12	23	35	46	56	59	<i>G</i>	1.05	1.06	36	38	14	10	2
50	1	3	10	19	28	37	41	<i>B</i>	1.31	0.95	19	45	22	14	0
75	0	2	5	11	19	25	28								
100	0	1	3	7	13	17	19								
150	0	0	2	3	7	11	12								
235	0	0	0	1	2	4	5								
255	0	0	0	0	1	3	3								

(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 is sufficient. Although, in general, each "inverse" intensity is 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} :

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

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. 2) 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 (on 0 to 255 scale) 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 to ϕ_B is shown in Fig. 5; this is a complementary mapping to the one from Fig. 3. 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. 5(b) presents γ_l computed under such a constraint and shown as intensity.

A crosstalk cancellation algorithm based on the non-negativity constraint is particularly simple. It requires 256×256 matrices γ_l and γ_r for each color component (such as in Fig. 5); operations other than table look-up are not needed. These 6 look-up tables computed for our stereoscopic setup are available from our web site [3]. 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. 4(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 $\mathcal{D}(\mathcal{T}^{-1}) = [0, 255] \times [0, 255] \setminus \Upsilon$.

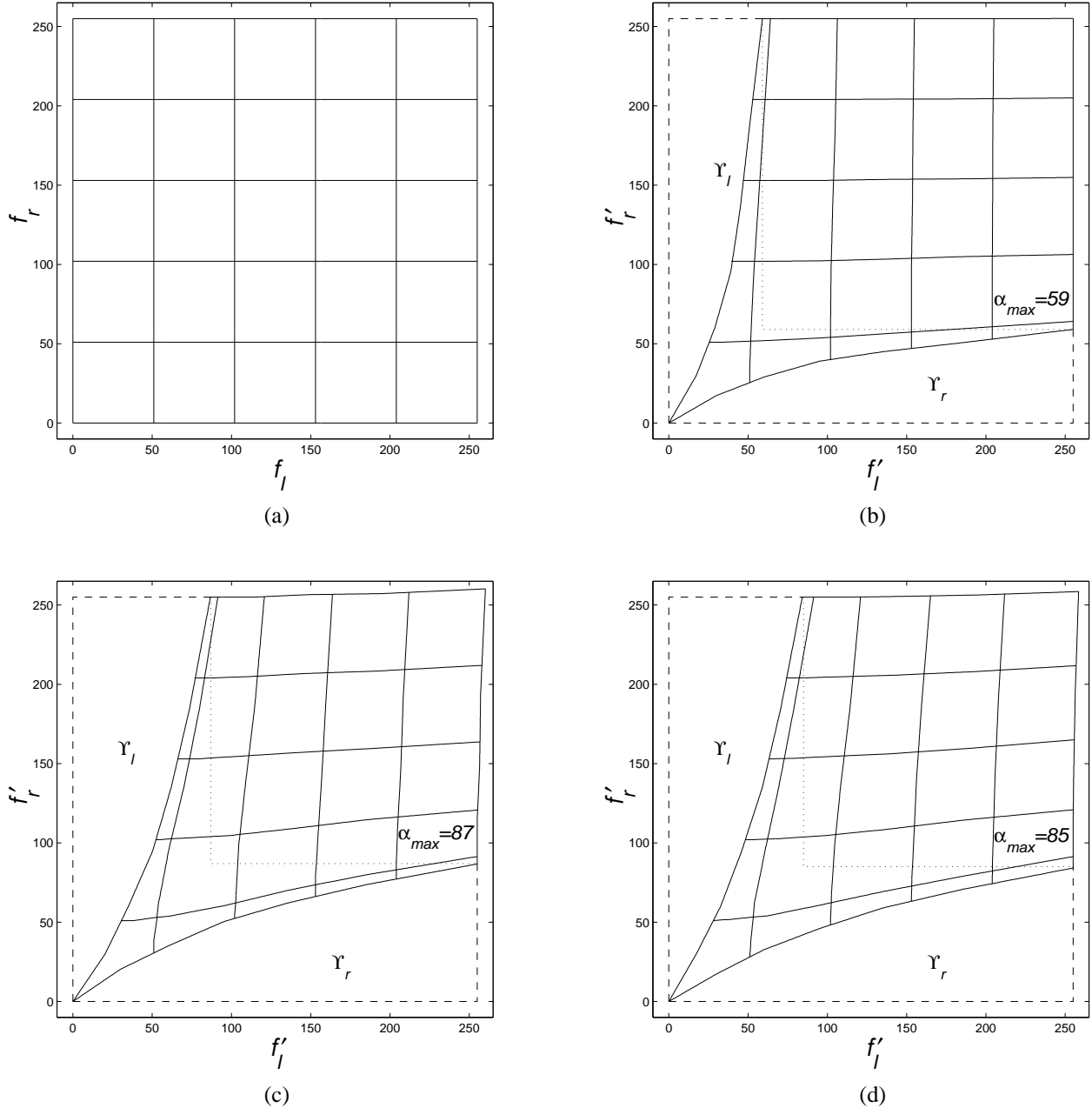


Fig. 4. (a) Regular grid of intensities $(f_l, f_r) \in \mathcal{D}(\mathcal{T})$ with levels 0, 51, 102, 153, 204, 255, and the same grid after mapping by \mathcal{T} (equation (2)) for (a) ϕ_R ; (b) ϕ_G ; (c) ϕ_B . See text for the meaning of Υ_l and Υ_r . Dotted lines show $\mathcal{S}(\alpha_{max})$.

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 $\mathcal{T}(0, f_r) = (\psi(f'_r), f'_r)$ and $\mathcal{T}(f_l, 0) = (f'_l, \psi(f'_l))$. These are boundary curves of the transformed grid in Figs. 4(b-d) that neighbor Υ_l and Υ_r , respectively. Also, let $\mathcal{S}(\alpha)$ be a variable-size square in the (f'_l, f'_r) plane defined by $[\alpha, 255] \times [\alpha, 255]$, where $0 \leq \alpha \leq \alpha_{max}$ and $\alpha_{max} = \psi(255)$ for each primary color (dotted square in Fig. 4). For $\alpha = \alpha_{max}$ 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 $[0, 255] \times [0, 255] \rightarrow \mathcal{S}(\alpha)$:

$$\begin{aligned} f_l &\leftarrow \alpha + f_l \frac{255 - \alpha}{255}, \\ f_r &\leftarrow \alpha + f_r \frac{255 - \alpha}{255}, \end{aligned}$$

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

$$\begin{aligned} f_l &\leftarrow \max(f_l, \alpha), \\ f_r &\leftarrow \max(f_r, \alpha). \end{aligned}$$

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.

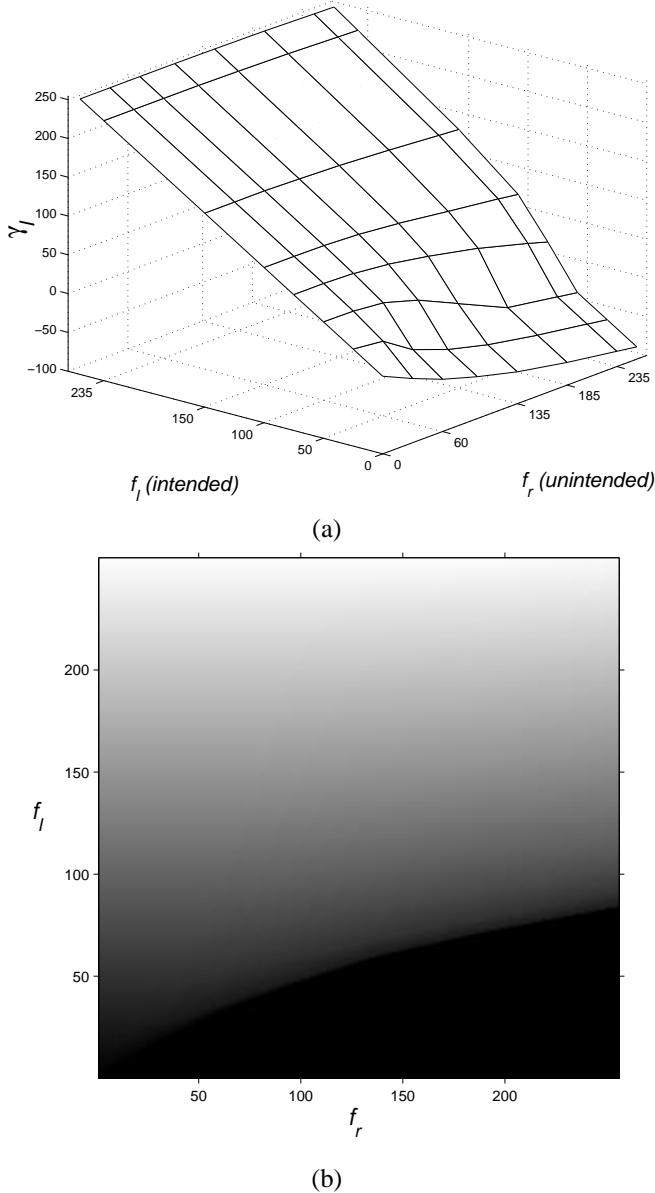


Fig. 5. B component of the left image, i.e., a solution γ_l to equation $f_l = \gamma_l + \phi_B(\gamma_r, \gamma_l)$ for $(f_l, f_r) \in [0, 255] \times [0, 255]$: (a) surface plot; (b) representation as intensity. Since T^{-1} is defined in $\mathcal{R}(T)$, the surface plot represents T^{-1} only for $\gamma_l \geq 0$. 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.

Note that for a fixed α the above mappings can be easily incorporated into the LUT discussed in Section IV 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).

V. EXPERIMENTAL RESULTS

We have applied the algorithms described above to the CCETT stereoscopic video sequences *piano*, *tunnel*, *train*, *manege*¹, to the NHK sequences *flowerpot* 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. 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 and NEC's *XP 17* 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 $\alpha = \alpha_{max}$ ($\alpha_{max} = 59, 87, 85$ 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 ghosting.

Then, we tested the non-linear mapping (saturation) also with $\alpha = \alpha_{max}$ 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 ghosting 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 $\alpha = 20-30$ 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*,

¹See Acknowledgments section.

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. The original and some processed images from the above list are available for inspection on our web site [3]).

We have found that for $\alpha=20-30$, the 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 were 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 unacceptable solution in quality-oriented applications.

Finally, we have tested the crosstalk compensation algorithm with the non-negativity constraint only (no pre-processing). With this approach although only partial crosstalk compensation was achieved, the 3-D perception was significantly improved. The cancellation was only slightly inferior to the one with non-linear mapping but it was clearly worse than for the one with linear mapping, both with $\alpha=20-30$. In addition to the fact that the method preserves image fidelity (only slight loss of dark-area detail) while largely suppressing the crosstalk, it also has the lowest computational complexity of all the algorithms tested. The approach may be of interest in applications where image fidelity is of primary concern.

To visually demonstrate how the compensation algorithm affects images, Figs. 6 and 7 show a small luminance window from field #0 of sequences *piano* and *train*, respectively. In *piano*, silver-white leaves are presented over a blue background, whereas in *train* a red-and-white semaphore is shown over a gray sky. In the middle rows, original images with crosstalk superposed digitally according to model (1) are shown; note slight ghosting on either side of the leaves. We have compared these images with the original images from Figs. 6(a-b) and 7(a-b) viewed through LC shutters using one eye at a time. For example, for the left eye, left monoscopic image with superposed model-based crosstalk was compared, by instantaneous switching, with the left original image that exhibited natural crosstalk due to leakage from the right image of the stereopair. In informal tests, the two images were judged almost identical. This test has confirmed the validity of our model (1). We have drawn another interesting conclusion from the above test. Since the crosstalk-prone areas in the tested images were close to image boundaries (leaves in *piano* and semaphore in *train*) and since the model was judged accurate, we can conclude that the assumption about spatial uniformity of the screen is not detrimental to the model and therefore to the crosstalk cancellation itself.

In the bottom row are shown images from Figs. 6(a-b) and

7(a-b) after the non-linear mapping with $\alpha=30$. Note the dark imprints in the background of either window corresponding to bright leaves and white semaphore sections 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. An extensive set of results (linear and non-linear mapping, various α 's, various images) can be found on our web site [3].

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.

VI. 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 require 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 inverse format conversion would be needed; a 400kB memory and a simple scaling or “max” operator on the board would suffice. 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 $\alpha=20-30$ 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 $\alpha=20-30$; higher values should be used only in extreme

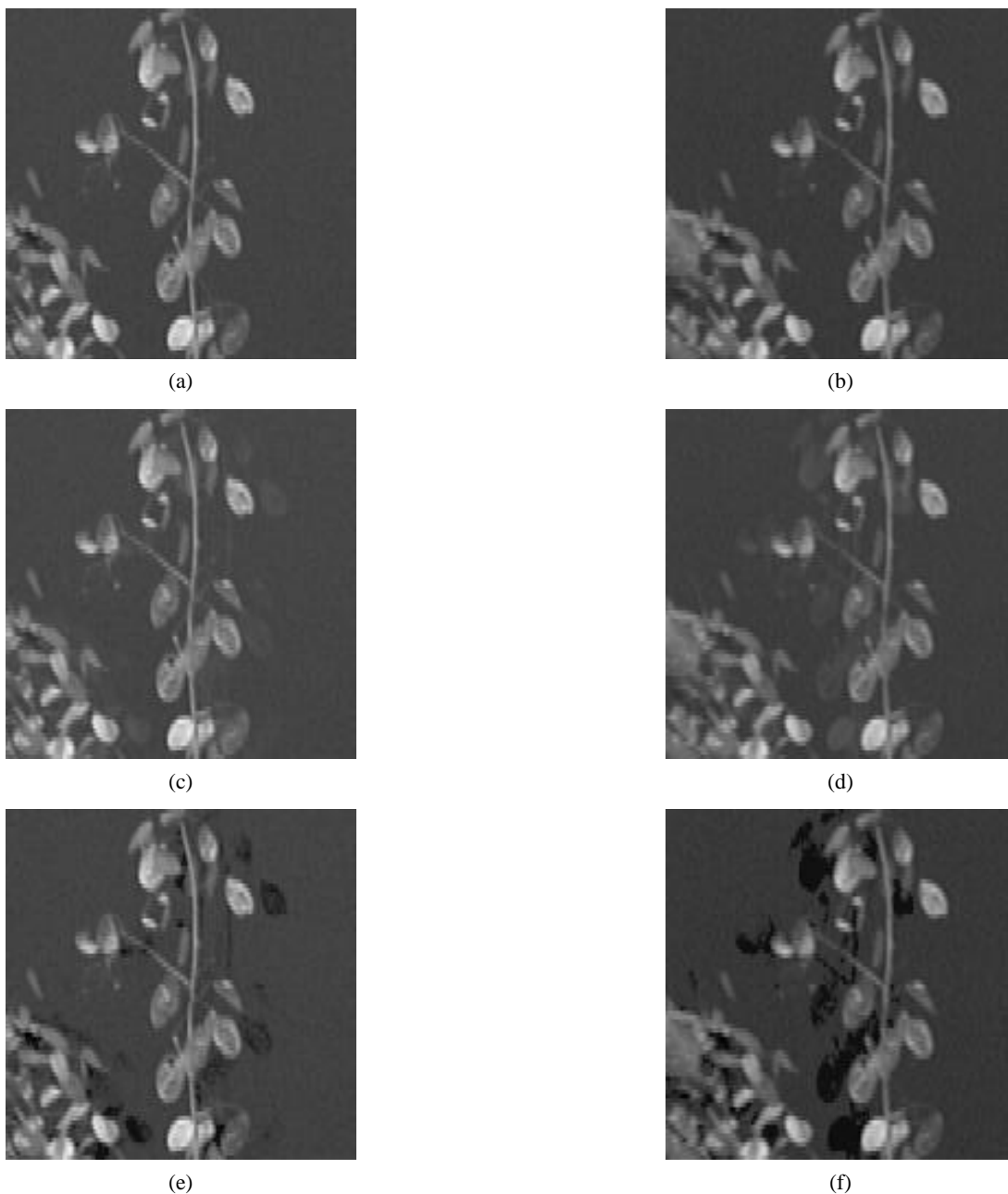


Fig. 6. Luminance of the original (a) left and (b) right fields #0 from *piano* sequence, (c-d) the same images with crosstalk superposed digitally according to model (2), and (e-f) original images compensated for crosstalk using non-linear mapping with $\alpha=30$. Since the sequence is 2:1 interlaced, vertical interpolation by 2 was applied to maintain correct aspect ratio.

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.

VII. 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, we validated model's accuracy 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

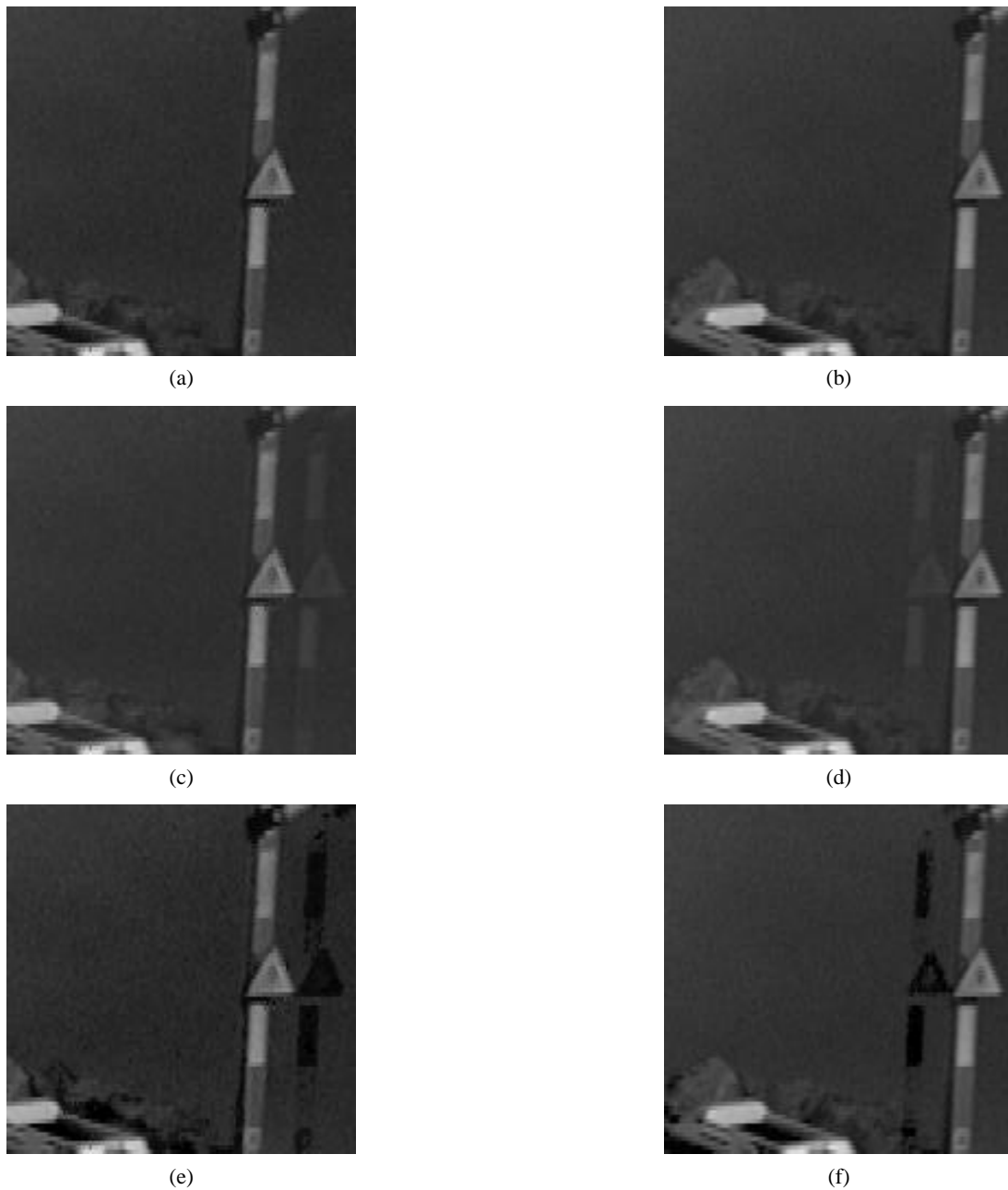


Fig. 7. Luminance of the original (a) left and (b) right fields #0 from *train* sequence, (c-d) the same images with crosstalk superposed digitally according to model (2), and (e-f) original images compensated for crosstalk using non-linear mapping with $\alpha=30$. Since the sequence is 2:1 interlaced, vertical interpolation by 2 was applied to maintain correct aspect ratio.

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. The proposed approach can be extended to handle the additional top- and bottom-screen crosstalk present in many systems; however additional measurements would have to be performed. We believe that the algorithm presented can be implemented today on advanced graphics boards, 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. We would also like to thank Dr. Albert Gołębowski and Mr. Anthony Mancini for their help in preparing some of the results.

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] <http://www.inrs-telecom.quebec.ca/users/viscom/english/publications>.
- [4] B. Lacotte, "Elimination of keystone and crosstalk effects in stereoscopic video," Tech. Rep. 95-31, INRS-Télécommunications, Dec. 1995.
- [5] J. Lipscomb and W. Wooten, "Reducing crosstalk between stereoscopic views," in *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems*, vol. 2177, pp. 92–96, Feb. 1994.
- [6] L. Lipton, "Compatibility issues and selection devices for stereoscopic television," *Signal Process., Image Commun.*, vol. 4, pp. 15–20, 1991.
- [7] L. Lipton, "True stereoscopic television: 3D-TV is feasible and striking," *Advanced Imaging*, pp. 28–30, July 1994.
- [8] I. Sexton, T. Bardsley, and A. Bhoopal, "Errors in depth," in *Int. Workshop on Stereoscopic and 3D Imaging*, pp. 235–242, Sept. 1995.
- [9] G. Street, "Method and apparatus for image enhancement." European Patent No. 819359, Feb. 1999.
- [10] W. Tam and L. Stelmach, "Stereo depth perception in a sample of young television viewers," in *Int. Workshop on Stereoscopic and 3D Imaging*, (Santorini, Greece), pp. 149–156, Sept. 1995.