ELIMINATION OF KEYSTONE AND CROSSTALK EFFECTS IN STEREOSCOPIC VIDEO

ELIMINATION OF KEYSTONE AND CROSSTALK EFFECTS IN STEREOSCOPIC VIDEO

Bertrand LACOTTE

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

22 décembre 1995

Rapport technique de l'INRS-Télécommunications no. 95-31

Summary

First of all, I would like to thank all the people who made this internship possible. Mr Gille Delisle, headmaster of the INRS-Télécommunications, Mr Janusz Konrad, in charge of my study, and Mr Henri Maitre, head of the Image option in the ENST. I also would like to thank all the students, professors and associates for their kindness and support. Finally, I thank all the Quebecers I met during this internship in Montreal because of their true friendship.

The aim of this internship was to study stereoscopic video, that is three dimensional television, first become familiar with this technology, and then improve the quality of the images. Two major defects were analysed during these five months. The keystone distortion, which is a geometric defect, and the crosstalk distortion coming from the technological limitations of display systems. This internship was all the more interesting as many laboratories troughout the world are studying three-dimensional advanced technologies, which might become a new standard in term of image viewing and broadcasting in the next century.

Contents

List of Figures

1 Introduction

After the High Definition Television (even if it is not yet available for the general public), the 3-dimensional television seems to be the next major step in the world of imaging, since it represents a move towards bifocal and thus natural vision. The availability ot the third dimension should enhance the telepresence of the viewer using the full capabilities of human visual perception. The principle is nevertheless not new, since the first 3D pictures appeared at the beginning of the century. With the evolution of technology in the field of image computing, the aim is now to study the best ways to create, record and watch stereoscopic videos, eliminate the possible distorsions and broadcast the images using the actual transmission standards. But one of the highest priorities remains to study the real impact of 3DTV on people watching such new images, since whereas the regular TV is well-known and adapted to the human observer, 3DTV standards still need to be defined. Numerous studies are driven throughout the world to break down and understand all the parameters concerning stereoscopic video.

2 3DTV Technology

2.1 The principle of stereoscopic images

Although a vivid perception of the surrounding environment can be obtained from a single eye, human vision is essentially a binocular process that transforms two images seen from slightly different viewpoints into a clear perception of solid 3D space. The fact that the two eyes can be stimulated separately to recreat the three-dimensional geometry is exploited by a number of two-dimensional stereoscopic display techniques. Depth perception is achieved by presenting to the left and right eyes images that are laterally displaced relative to one another (this lateral shift is named parallax). These images stimulate the disparate perception that is normally produced when horizontally separated eyes look at objects in 3D space, thus creating three-dimensional vision. Techniques such as holography with a viewing angle of 360 degrees are obviously out of our concern since stereoscopic vision normally allows very small viewing angles of the watched objects.

2.2 Parallax

When looking at an object in the real world, the eyes both converge and focus on it. Thus, for the left and right images perceived by the eyes, there is no lateral shift (no parallax or zero parallax) for the object on which the eyes are converged. However, for the objects behind, there is a positive shift (positive parallax) between the left and right images ; the further away the objects, the larger the shift. For the objets in front, there is a negative shift (negative parallax) ; the closer the objects to the viewer, the larger the negative shift. When a stereoscopic video sequence is displayed

on a screen, the objects without parallax (no shift) appear in the plane of the screen, the ones with a positive parallax seem to be inside the monitor, and the ones with a negative parallax seem to appear in the space between the screen and the viewer.

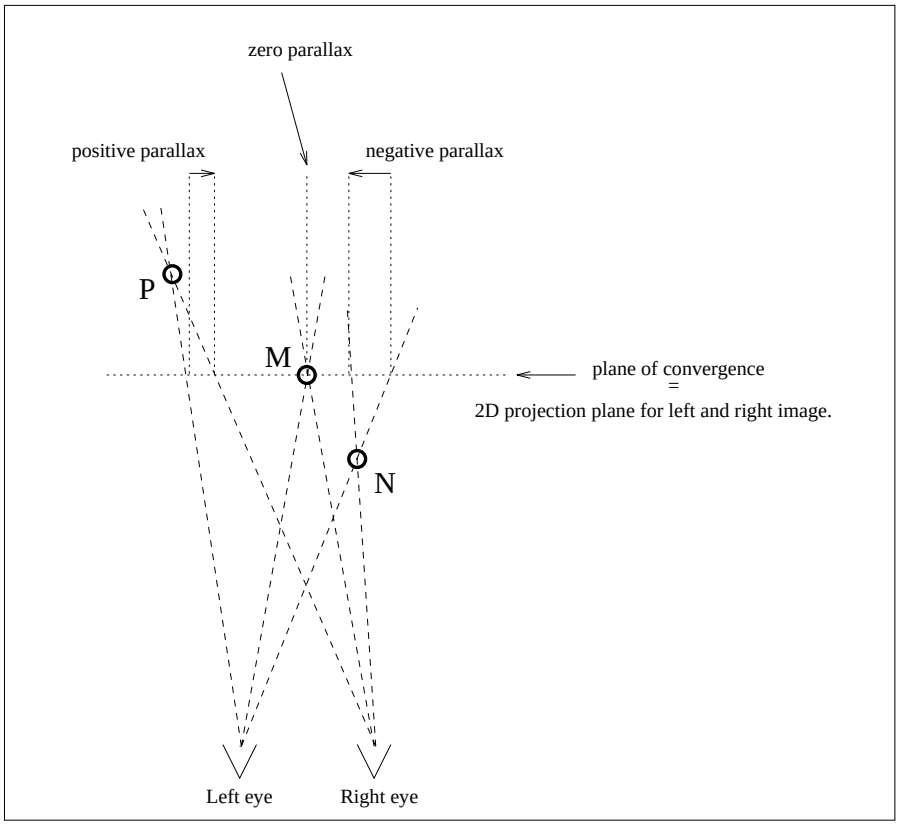


Figure 1: Positive, zero, and negative parallax in 3D scenes

2.3 Pick-up devices and geometry

A stereoscopic system consists of a pair of video cameras mounted side by side to obtain left and right images. Actually, there are two types of camera setup with respect to the convergence : parallel cameras and toed-in cameras.

Since some symbols will be cited quite frequently in the following, we itemize them before going on with greater details.

t	Camera separation, the distance between the two optical centers.
f	Focal length
w	Sensor width
2ϖ	Horizontal angle of view of the camera.
M	Frame magnification, ratio of screen width to sensor width.
P	Image parallax
C_l, C_r	Optical centers of cameras.
$O_{sl}O_{sr}$	Origins of the sensor planes.
2ϕ	Angle of convergence produced by the optical axes.
V	Viewing distance.
h	Sensor axial offset.

2.3.1 Parallel cameras

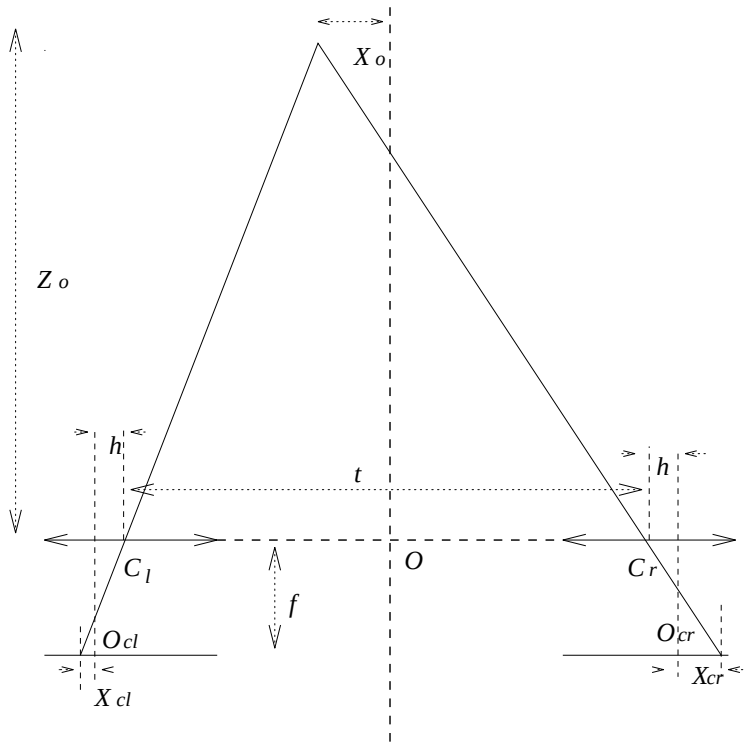


Figure 2: Parallel configuration of cameras(in horizontal plane)

The convergence is ensured by the external shift h of each sensor with respect to the optical axis of its lens (we are in the situation of parallel axes, nevertheless we need a convergence plane). The transformation of a real 3D point (X_o, Y_o, Z_o) into a left (X_{cl}, Y_{cl}) 2D-point and a right (X_{cr}, Y_{cr}) 2D-point is given by the following equations :

$$X_{cl} = \frac{f(t + 2X_o)}{2Z_o} - h, \tag{1}$$

$$X_{cr} = -\frac{f(t-2X_o)}{2Z_o} + h \quad (2)$$

$$Y_{cl} = Y_{cr} = \frac{Y_o f}{Z_o} \quad (3)$$

2.3.2 Toed-in cameras

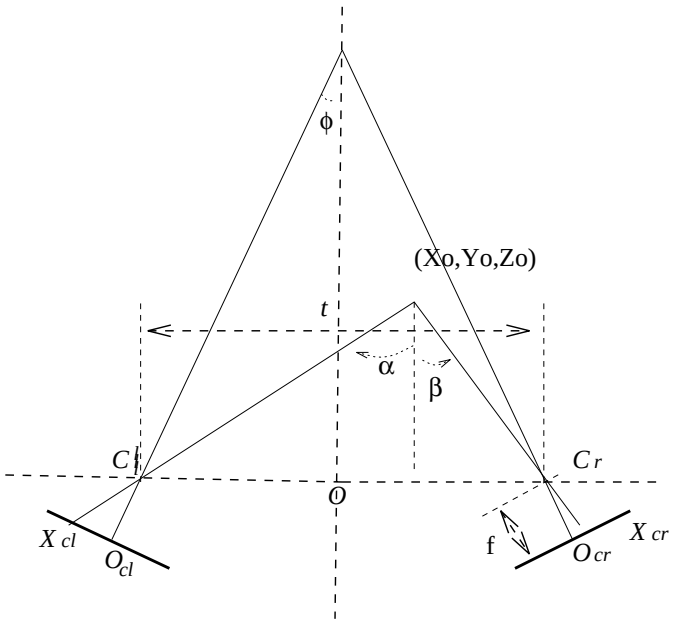


Figure 3: Toed-in configuration of cameras (in horizontal plane)

The situation of a toed-in camera setup is illustrated in Figure 3. The coordinates for the two sensors are given by Woods, Docherty and Koch [4]. To simplify the expressions we note that

$$\alpha = \arctan\frac{t+2X_o}{2Z_o}, \qquad \beta = \arctan\frac{t-2X_o}{2Z_o}.$$

To obtain the relationship between the object coordinates and the sensor coordinates, we refer to Figure 4.

The computation for determining X_{cl}, X_{cr} is direct because $\angle ICO_l = \angle FC_lG = \alpha - \phi$. Notice that

$$\frac{Y_{cl}}{Y_o} = \frac{IC}{CG}$$

and

$$CG = \frac{Z_o}{\cos\alpha}.$$

Thus, we can write:

$$X_{cl} = f\tan(\alpha - \phi), \quad (4)$$

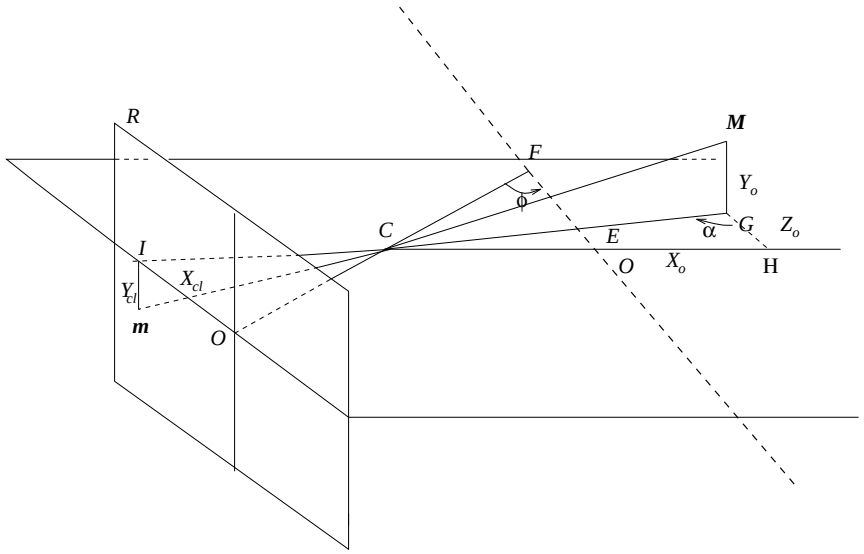


Figure 4: Computing the sensor coordinates on the left sensor $CH = X_o, MG = Y_o, GH = Z_o, \angle CEO = \angle CGH = \alpha$

$$X_{cr} = f \tan(\beta - \phi) \quad (5)$$

$$Y_{cl} = \frac{f Y_o \cos\alpha}{Z_O \cos(\alpha - \phi)} \quad (6)$$

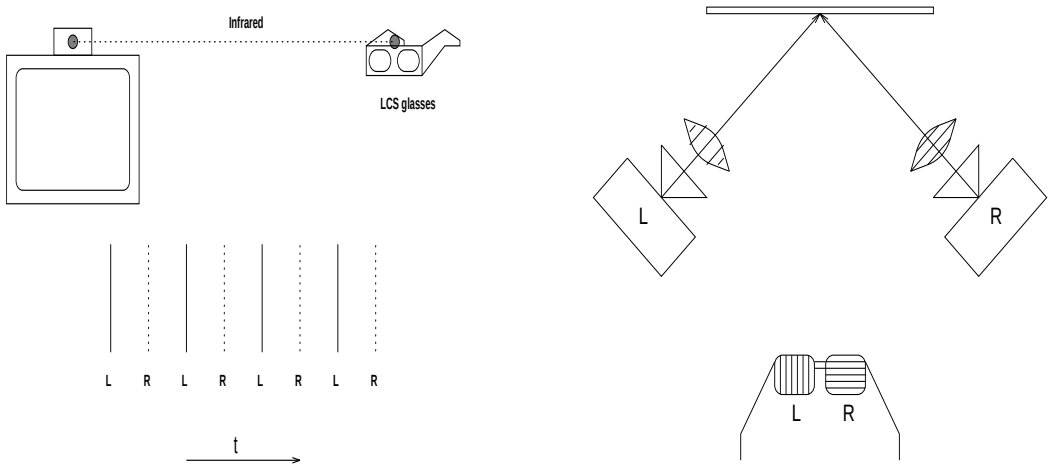
$$Y_{cr} = \frac{f Y_o \cos\beta}{Z_O \cos(\beta - \phi)} \quad (7)$$

2.4 3DTV displays

2.4.1 Displays using glasses

At present, in a stereoscopic system the most convincing picture quality can be obtained by using two displays (e.g. two TV projectors) to view the left and the right images followed by a view separation using polarizing filters. This method requires a "metallic" screen so as to preserve the polarizations of the projectors, and could be applied in medium-size theaters in the near future. An alternative display method is the time-sequential display technique, which only requires one display device. View separation is obtained by active shutter glasses, mostly of Liquid Crystal Shutter (LCS) type, such as the CrystalEyes used at the INRS-Télécommunications) which are synchronized to the alternating views on the cathode ray tube by infrared links. Another solution is to mount an electronically switched polarizing filter plate in front of the display. In this case the user has to wear only passive polarizing glasses as for the dual display method.

A time-sequential system such as this should have at least twice the standard frame rate to avoid flicker. The main problem in both methods is the acceptance of viewing glasses. 3DTV is jugded to be very attractive, but many people accept the



time-sequential display, LCS glasses polarized projectors and glasses

Figure 5: Two of the most used stereoscopic display devices

glasses only for a short time.

A special version of 3DTV displays is the helmet-mounted configuration used in robotics and recently in the new field of interactive TV to create "Virtual Reality". In principle, two very small monitors are applied directly to the eyes (e.g. LCD displays). But up to now the resolution is too low, especially for the wide-angle views wanted in these applications.

2.4.2 Autostereoscopic 3DTV displays

A breakthrough in 3DTV, especially in the consumer mass market, will be the availability of display technologies which do not need viewing glasses. But also in some professional applications, e.g., in medecine for surgical operations or in aircraft cockpits, glasses may not be acceptable. The most popular principle is the lenticular lenses method. Small narrow-spaced cylindrical lenses are mounted in front of a picture which consists of stripes of alternating views for the left and right eye. The lenses are such of shape and configuration that each eye can only see its corresponding view. Using a higher number of views, horizontal look-around capability of a scene can be obtained. TV applications of this method are working with a single cathode ray tube (with high resolution to obtain sufficient resolution for each view), or with mutiple TV projectors. Similar approch is a barrier system which masks the unwanted view to each eye, e.g., in a sandwich arrangement of an LCD image display and an LCD barrier plate. Further methods are parallax illumination systems which consist of a transparent liquid crystal image source and an illumination plate as backlight. The illumination plate sends out, in time-sequential mode, its directional light via the transparency into the eyes of the viewer. Finally, the volume-scanning systems should be mentionned. A recent development works with an ultrasonic deflected laser

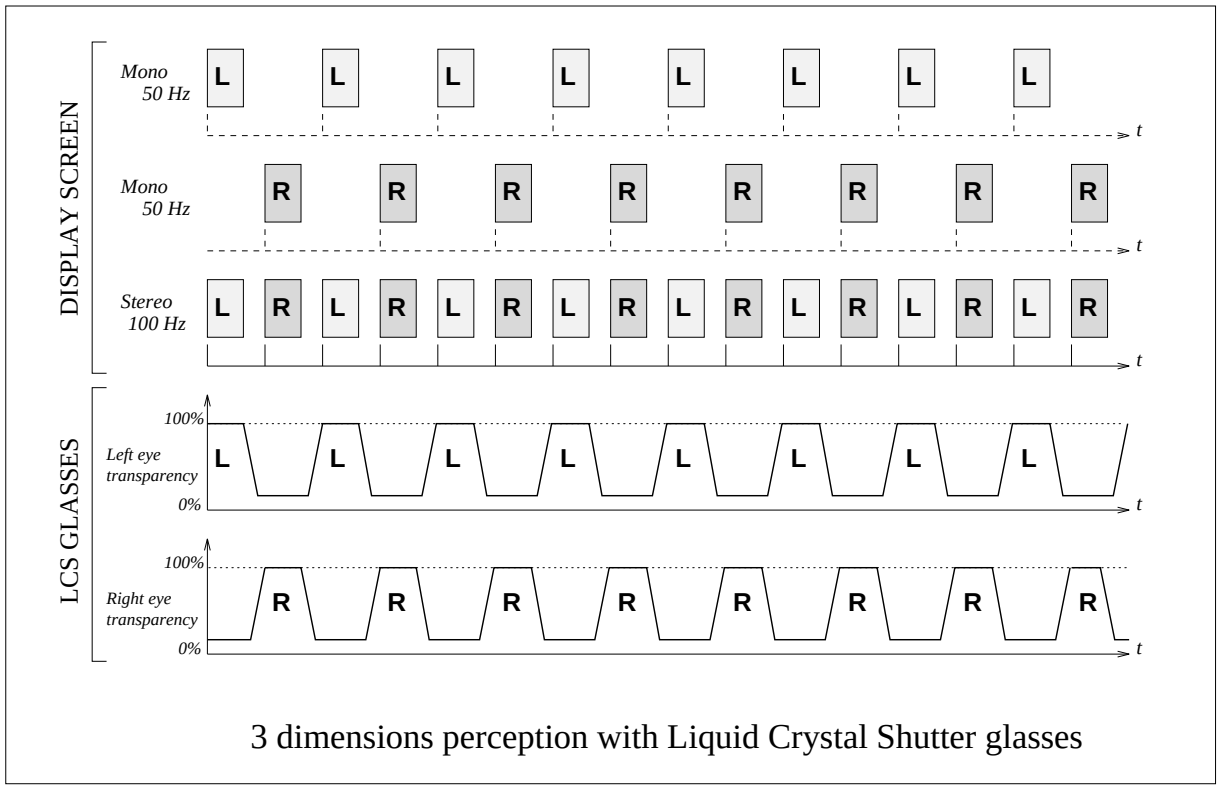


Figure 6: Time-sequential display device with LCS glasses

beam projected onto a rotating screen. The 3D image appears in a cylindrical space. Another volumetric display uses a variofocal mirror which, varying with sufficiently high frequency, produces a TV monitor picture within a cube.

3 Geometric distortions in stereoscopic systems

Stereoscopic distortions are ways in which a stereoscopic image of a scene differs from actually viewing the scene directly. There are a number of different types of image distortions in stereoscopic video systems, coming from the geometry of the pick up and display devices, but also from the optical equipment. Human factors also have to be taken into account. We will discuss these distortions, and especially the keystone distortion (in the toed-in camera configuration).

3.1 Main distortions

3.1.1 Depth plane curvature

Contrary to the parallel camera configuration, the toed-in camera configuration results in a curvature of the depth planes. This will result in objects at corners of the image appearing further away from the viewer than objects at the center of the im-

age. In contrast, the parallel camera configuration results in depth planes which are parallel to the surface of the screen. Depth plane curvature is closely linked with the keystone distortion, which will be discussed later. The depth plane curvature could lead to wrongly perceived relative object distances on the display and also disturbing image motions during panning of the camera system.

3.1.2 Depth non-linearity

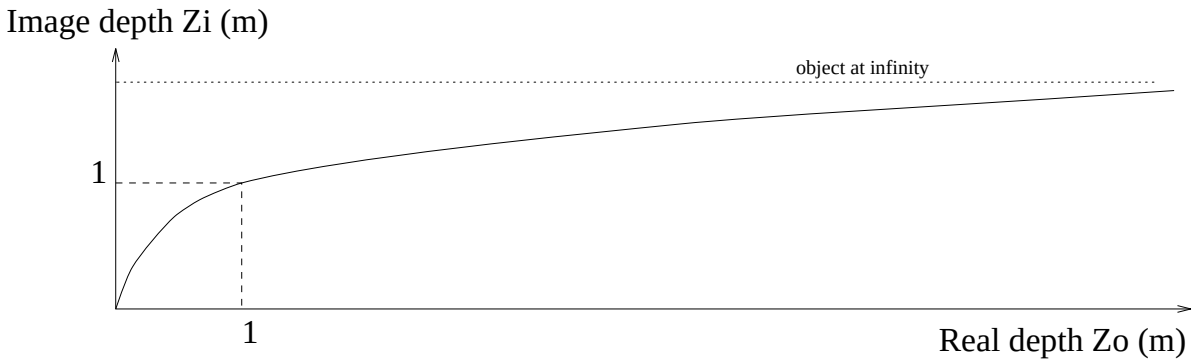


Figure 7: Depth non-linearity for convergence and viewing distances at 1 meter

It can be seen from Figure 7 that the depth is stretched between the viewer and the monitor and compressed between the monitor and infinity. This non-linearity of depth on the display can lead to wrongly perceived depth on the screen and if the camera system is in motion, it can lead to false estimation of velocity. At first the vehicle will appear to be approaching the structure rather slowly but once the structure comes closer to the camera than the convergence distance, the vehicle will appear to accelerate. A linear relationship between image depth and object depth can only be obtained by configuring the stereoscopic video system such that infinity is displayed at image infinity on the stereoscopic display. The viewing distance (while facing the screen) also directly interfere with depth perception.

3.1.3 Shear distortion

A disadvantage of binocular stereoscopic displays is that the stereoscopic image appears to follow the observer when he/she changes viewing position. A sideways movement leads to a "shear distortion" resulting in a sideways shear of the stereoscopic image about the surface of the monitor. Images out of the screen will appear to shear in the direction of the observer whereas images behind the surface of the screen shear in the opposite direction. Actually, observer motion will also lead to false perception of motion in the image.

3.1.4 Depth and size magnification

An analysis of image magnification reveals that there can be a mismatch between depth magnification and size (width and height) magnification. This is particularly so when there is a non-linear relationship between image on object depth. This leads to an image appearing flat or conversely stretched.

3.1.5 Lens distortion

Lens radial distortion, often called pin-cushion or barrel distortion, is another source of image distortion and induced vertical parallax. It is caused by the use of spherical lens element, resulting in the lens having different focal lengths at various distances from the center of the lens. Among common lenses, radial distortion is worst for short focal length lenses. Aspherical lenses should be used especially when short focal length lenses are required for wide angle shots.

3.1.6 Other interferring human factors

The complexity of highly advanced communication systems like 3DTV, with regard to realization and economy, is only to be solved by the knowledge of its psychovisual factors. Considerable work has been done already in connection with human factors' evolution for HDTV, but many questions are still open or need further investigation. For example, small 3D pictures create a puppet theater effect whereas very large, panoramic pictures seem to allow to dispense true 3D, filling completely the viewers field of vision. In HDTV research it was found that the overall psychological impact of stereoscopic image equals that of flat 2D images twice their size. But also, how important is the "look around capability" ? Must the reproduction of perspective relations be true ? Experience in robotics showed that an exaggerated perspective is often more propitious and in entertainment the spectators often prefer "pleasant pictures" to correct ones. There are also strictly technological problems, one of these named crosstalk. When using a single monitor, left and right images are to be displayed alternatively on the screen at a high speed (more than 100Hz to avoid flicker). But the phosphors used on the screen have a too long relaxation time (depending on the color too). Also due to the imperfect active LCS glasses (opaque LCS still passes about 10% of light through), left and right eyes can partially see the wrong image. This results in "ghosts" due to the horizontal parallax, creating a bad visual effect,and disturbing the stereoscopic perception. Many questions in human factors research also arise in connection with data reduction for signal transmission. The phenomenon of subjective image quality enhancement in stereoscopic viewing has to be investigated carefully in connection with possible irrelevance coding schemes. Can one of the two views be displayed with a lower quality ? And how is it related to the effect of "suppression" and the "leading eye" ? Are there further phenomena that could affect 3D imaging ? Furthermore, the "3D production grammar" for comfortable 3D viewing without eye-strain needs to be completed.

As for the eye-strain, two main factors are to be taken into account : accommodation and vergence, and vertical parallax. A widely discussed limitation of field-sequential stereoscopic displays is the association between accommodation and vergence. In real world viewing vergence and accomodation are normally closely linked visual actions, whereas stereoscopic displays require a different visual action. The eyes must remain focused at the surface of the screen at all time regardless of where the eyes are verged in the stereo monitor. Excessive screen parallax can lead to stereoscopic images appearing out of focus and/or the viewer being unable to fuse the images. This can be due to the association between vergence and accomodation. But the responses are very different, depending on the viewers. As for vertical parallax, which is nothing but simple distorsion coming from the lens or the keystone effect, it was mesured that homologous points should have less than 7mm of vertical parallax for image fusion to be possible. Eye strain was apparent at higher values of the vertical parallax. Needless to say, vertical parallax should be reduced as much as possible to produce an easily viewed image. This is one of the purposes of the keystone distorsion rectification to be discussed now.

3.2 Keystone distortion

3.2.1 Description

A well known effect of the toed-in camera configuration is the keystone distortion. Keystone distortion causes vertical parallax in the stereoscopic image due to the imaging sonsors of the two cameras being located in different planes. The effect of keystone distortion upon the display of a grid located at the camera convergence distance is shown figure 8.

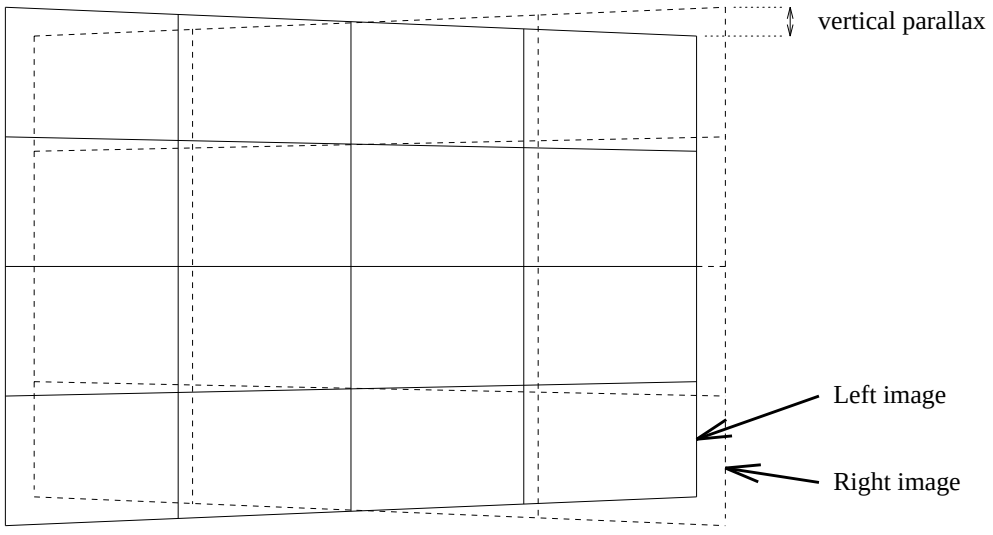


Figure 8: vertical parallax caused by keystone distorsion

In one of the cameras the grid appears larger at one side than at the other. In

the other camera the effect is reversed. This results in a vertical difference between homologous points which is called vertical parallax. The amount of vertical parallax is greatest in the corners of the image and increases with increased camera separation, decreased convergence distance and decreased focal length (for example : a lens with a focal length of 3.5mm, a convergence distance of 1m, and a camera separation of 75mm, would exhibit vertical parallax of 8.2mm in the corner of a 40cm screen). It can be seen from the figure that horizontal parallax is also induced. This is the source of the depth plane curvature mentioned earlier. The parallel camera configuration does not exhibit keystone distorsion.

3.2.2 Epipolar constraint

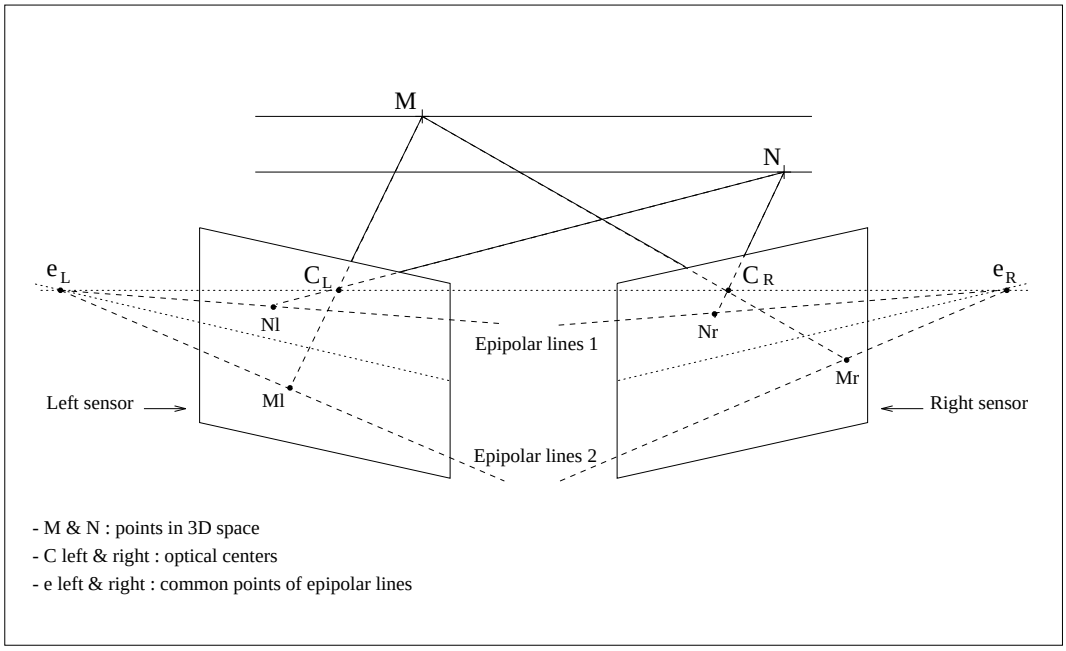


Figure 9: The epipolar lines principle

The epipolar constraint consists of the following geometric relationship : if e_l and e_r are the intersection points of the base line (determined by the left and right optical centers) with the left and right sensors, any horizontal line in the 3D space parallel with the base line should have its images going through e_l and e_r. This explains the deformation of a grid placed in the convergence plane. Thus any point in the plane determined by $M, C_l and C_r$ will have its images located on the on the epipolar lines linked with M. Another important result is due to this epipolar constraint : if we take any point on a sensor image of a point M in the 3D space, we know on which epipolar line it is (e_l & e_r can easily be computed), so we also know that its corresponding point (for the same 3D point M) on the other sensor is located on the symetric epipolar line. This will be useful in the computation of the unknown convergence angle in stereoscopic sequences to eliminate the keystone effect.

3.2.3 Rectification for a known convergence angle

The rectification technique we are going to use consists of projecting the sensor images on a plane parallel to the base line of the cameras. This algebraic transformation is made of a rotation and a translation. The rotation allows the new planes to be in the same plane parallel to the base line of the cameras; the translation adjusts the width in moving the two planes to the new determined positions.

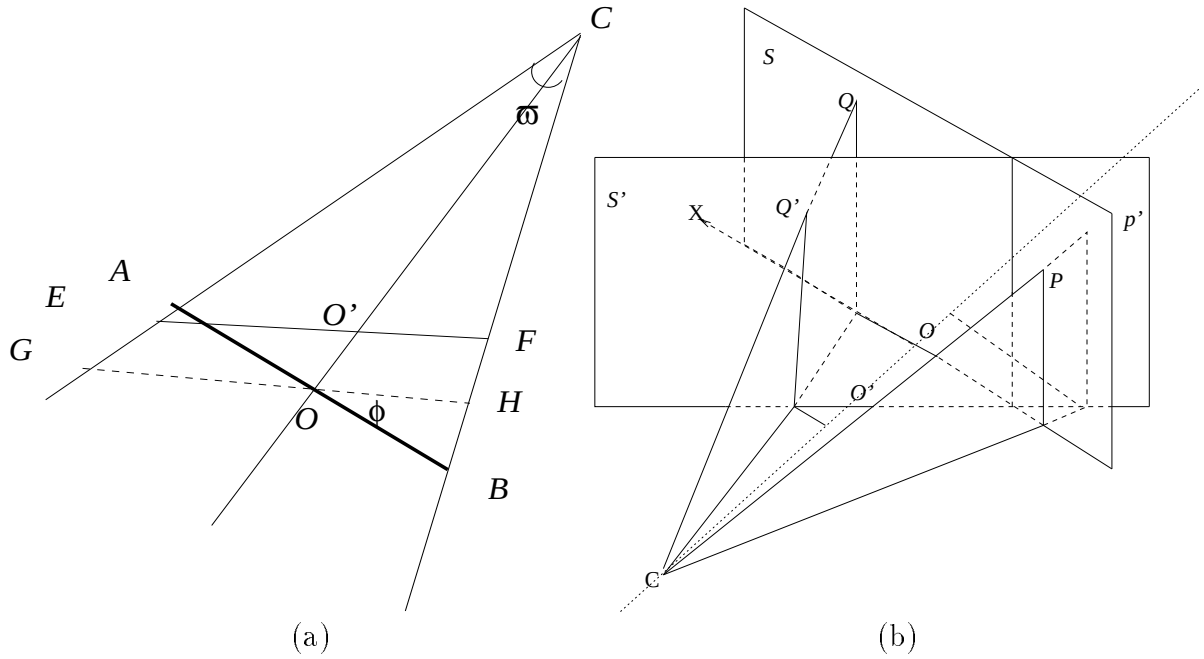


Figure 10: Determining the position of the rectification plane

Let w be the width of the sensor Figure10, and 2ϖ the angle of view of the camera. They obey the relationship : $\varpi = \arctan \frac{w}{2f}$. The final plane of rectification and the optical axis CO intersect at O'. The position of the plane is characterized by length of $f' = O'C$.

$$OG = \frac{w \cos \varpi}{2 \cos(\phi + \varpi)}, OH = \frac{w \cos \varpi}{2 \cos(\phi - \varpi)} GH = \frac{w}{2} \cos \varpi \left(\frac{1}{\cos(\varpi + \phi)} + \frac{1}{\cos(\varpi - \phi)}\right)$$

The width of the image projected in the plane of rectification (EF) is equal to w. Therefore,

$$f' = \frac{2f}{\cos \varpi \left(\frac{1}{\cos(\varpi+\phi)} + \frac{1}{\cos(\varpi-\phi)}\right)}. \tag{8}$$

Assume that the old sensor plane is S. The plane of rectification is S'. The angle between the two planes is ϕ (in the anti-clockwise direction) as showed in Figure 10. We always have:

$$\frac{y^r}{y} = \frac{x^r \cos \phi}{x} = \frac{f' - x^r \sin \phi}{f}$$

This induces

$$x^r = \frac{f'}{f\cos\phi + x\sin\phi}x \tag{9}$$

$$y^r = \frac{f'\cos\phi}{f\cos\phi + x\sin\phi}y \tag{10}$$

. Alternatively, if the rotation is in the clockwise direction, the relationship is similar:

$$x^r = \frac{f'}{f\cos\phi - x\sin\phi}x \tag{11}$$

$$y^r = \frac{f'\cos\phi}{f\cos\phi - x\sin\phi}y \tag{12}$$

This system is equivalent to a parallel system with the same optical center and a fictitious focal length $f^r = f'\cos\phi$. Therefore, there will be no longer vertical parallax. We can compute the new coordinates Y^R_{cl}, Y^R_{cr} directly from the formula given for parallel cameras, or by substituting Y^R_{cl}, Y^R_{cr} by the relationships given above.

$$Y^R_{cl} = \frac{Y_o f'\cos\phi}{Z_o}$$

Obviously the image center O'_{sl} will not be the center of the new image. The new coordinates of the image border $X^1_{sl} = -\frac{w}{2}$ and $X^2_{sl} = \frac{w}{2}$ can be computed:

$$\begin{aligned} X^1_{cl} &= -\frac{f'w}{2(f\cos\phi - \frac{w}{2}\sin\phi)} \\ X^2_{cl} &= \frac{f'w}{2(f\cos\phi + \frac{w}{2}\sin\phi)} \\ X^1_{cl} + X^2_{cl} &= -\frac{f'w^2\tan\phi}{2f^2\cos\phi(1+\tan\varpi\tan\phi)(1-\tan\varpi\tan\phi)} \end{aligned}$$

The fictitious system will have an equivalent lateral shift h^r, it defined by the position of the plane, focal length, the width of the sensor and the angle between the sensor plane and the plane of rectification.

$$h^r = f'\cos\phi + \frac{f'w^2\tan\phi}{4f^2\cos\phi(1+\tan\varpi\tan\phi)(1-\tan\varpi\tan\phi)} \tag{13}$$

From the equation (8), we can conclude that a necessary condition for rectification is

$$\phi + \varpi \neq \frac{\pi}{2}.$$

3.2.4 Keystone rectification implementation

The biggest problem involved in the rectification is how to move the pixels from the original plane into the rectified one, since after the mathematical transformation, new pixel locations will not fall onto the sampling grid. Actually, we are going to do exactly the reverse. After calibrating the rectified plane, each pixel from this plan will be transposed in the "real" plan, and then its luminance and colour will be calculated using interpolation (e.g., separable cubic interpolation).

The program for keystone rectification will execute the following instructions:

- initialize the variables,
- open a stereoscopic sequence,
- collect the informations (angle, width of sensor,...),
- repeat twice (for the left and then the right image sequences with the appropriate changes) :
  - compute the coordinates of the high-left and low-right corners of the rectified images,
  - repeat for each image from the sequence :
    - load the original image,
      - repeat for each pixel from the rectified plan (line by line) :
      - - search the corresponding pixel in the non-rectified plan,
      - - interpolate its components (R,G,B or Y,Cr,Cb),
      - - store the new values in the rectified image,
      - save the new image and close the old one,
- close the sequence,
- clean memory.

3.2.5 Results and comments

Before going further into the results, we first describe the interpolation function used in the program. The cubic convolution interpolation used in the program is more accurate than the nearest-neighbor algorithm or linear interpolation method. Although not as accurate as a cubic spline approximation, cubic convolution interpolation can be performed much more efficiently.
If (x, y) are the coordinates of a pixel after the geometric transformation located within the rectangle $[x_j, x_{j+1}] \times [y_k, y_{k+1}]$ delimited by four pixels, for any component R, G, B or Y, C1, C2 the value of the component at the position (x, y) is given by the following function g :

$$g(x,y) = \sum_{l=-1}^{2} \sum_{m=-1}^{2} c_{j+l,k+m} u\left(\frac{x - x_{j+l}}{h_x}\right) u\left(\frac{y - y_{k+m}}{h_y}\right) \tag{14}$$

where u is the interpolation kernel :

$$u(s) = \begin{cases} \frac{3}{2}|s|^3 - \frac{5}{2}|s|^2 + 1 & 0 < |s| < 1 \\ -\frac{1}{2}|s|^3 + \frac{5}{2}|s|^2 - 4|s| + 2 & 0 < |s| < 2 \\ 0 & 2 < |s| \end{cases} \tag{15}$$

and $c_{j,k}$ is the component value for the pixel (x_j, y_k). This interpolation allowed a very precise and accurate geometric transformation of the images, with practically no loss in quality, contrary to the nearest-neighbor algorithm which gave really bad results, creating clearly visible step effects.

This rectification of keystone effect was applied on the sequences available at the INRS-télécommunications, but unfortunately those sequences were acquired with a very little convergence angle (convergence distance was more than 3 meters), so that the keystone effect had no real impact on the images. At most the rectification consisted of a 2-pixel vertical shift in the corners of the 40cm screen. The viewer could not say there was a real improvement between the original and the rectified image. But it is clear that for little convergence distance (less than 1 meter) the impact of such a rectification would be a real plus in 3D perception and would supress part of the discomfort coming from this unwanted vertical parallax.

3.2.6 Determining an unknown angle for toed-in cameras

The stereoscopic sequences we were using were precisely described so that we knew the convergence angle, but for most of the stereoscopic images, we have no information about the way they were acquired.

Because of the fact that the rectification seemed to be irrelevant for the used sequences, this part was not furthermore studied. Nevertheless we give the principle for finding the convergence angle.

For each point m_l in the left image, we know that the correponding point m_r in the right image is on the epipolar line determined by m_l and the angle of convergence ϕ. Given a value for ϕ, the idea is to search on this epipolar line for the point which is the most likely homologue of m_l by using a measure of similarity $F(m_l, m_r)$ between the two points, for instance luminance, and/or chrominance similarity. m_r is the point that minimizes $F(m_l, m_r)$. The sum of the values of F for all points in the left image will get smaller as ϕ is closer to the true convergence angle. Thus the optimal searched convergence angle will satisfy :

$$\phi^* = arg \min_{\phi} \sum_{m_l} F(m_l, m_r(\phi)) \tag{16}$$

left image

right image

Figure 11: Original stereoscopic sequence

left image

right image

Figure 12: Rectified images for a hypothetical convergence angle

Note that for each m_l, m_r is found by a one dimensional search (epipolar constraint) This method could allow to rectify any stereoscopic sequence, or at least the ones with a wide convergence angle creating a real discomfort when watching them.

4 Crosstalk effect

4.1 View separation

As mentioned before, the aim of stereoscopic displays is to show to each eye a different picture such as in real 3D vision. This is in contrast with the two-dimensional viewing where both eyes see the same image and thus no 3D effect is observed. In the field of "home stereoscopic video devices", the separation of left and right images is not that easy. Although few electronic firms are working on autostereoscopic displays, most devices consist of a screen on which left and right images are alternatively displayed at twice the standard frame rate to avoid flicker (see Figure 6). The viewer has to wear LCS glasses triggered by the vertical synchronization signal to allow each eye to see its own image and not the one of the other eye.

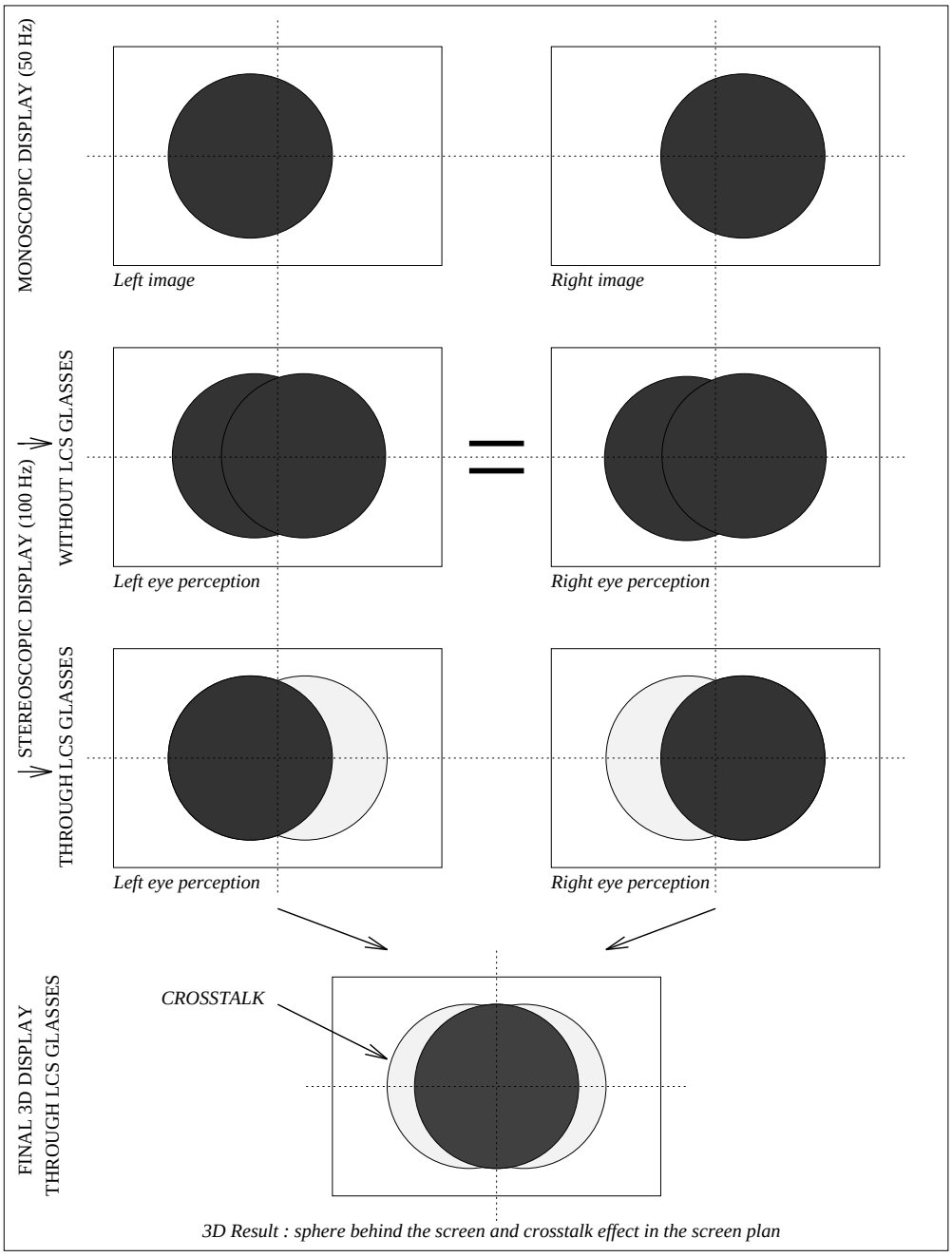


Figure 13: Crosstalk effect

Unfortunately, due to the limitations of the current technology in display devices, the separation of left and right image is not perfect. There are two major problems. Firstly, the monitor is used at a high frequency (100 to 120 Hz), so that the phosphors on the screen do not have enough time to come back to their lowest state of energy between the left and right image display. Secondly, the LC shutters of the glasses cannot close down to 0%. Part of the light can go throw the opaque LC shutters, allowing an eye to partially see the other image. Both effects result in each eye seeing its own image, plus a superposition of the image aimed at the other eye. Since the main difference between the left and right image is the horizontal shift (parallax), the left eye sees the left image, plus a "ghost" of itself, which is the shadowed right image (same for the right eye). This defect is called crosstalk effect. When watching a stereoscopic video sequence with crosstalk effect, the viewer sees the 3D image, but also echos and shadows on each side of objects, especially when they are bright and in front of a dark background. This can be enough to prevent the eyes from seeing clear 3D images, or at least represents a real discomfort when watching them.

4.2 Crosstalk elimination

4.2.1 Measuring the crosstalk

After a few psychovisual experiments, we came to the conclusion that the crosstalk had to be studied component by component (e.g. red, green and blue), to match the way the images are displayed on the screen. The first results showed us that there was no simple link between the image creating crosstalk and the generated "echo" in the other image.

Each pixel in the picture is decomposed into three integer components red (R), green (G) and blue (B), each one with an intensity between 0 and 255. We decided to measure the impact of the right picture on the left one for each R,G, and B component. If φ is the crosstalk function, we have assumed the following model for the crosstalk (until the end, we will only consider the green component) :

$$G_{left} + \varphi(G_{left}, G_{right}) = G'_{left} \tag{17}$$

$$G_{right} + \varphi(G_{right}, G_{left}) = G'_{right} \tag{18}$$

.

Above, G is the original green component value and G' is the target component value to be perceived by the viewer after the superposition of G_{right} on G_{left}. First, we need to evaluate φ.

4.2.2 Psychovisual evaluation of φ

In order to measure the crosstalk, we used the following experiment. The screen was divided into two parts. The viewer was watching the image only with his left eye

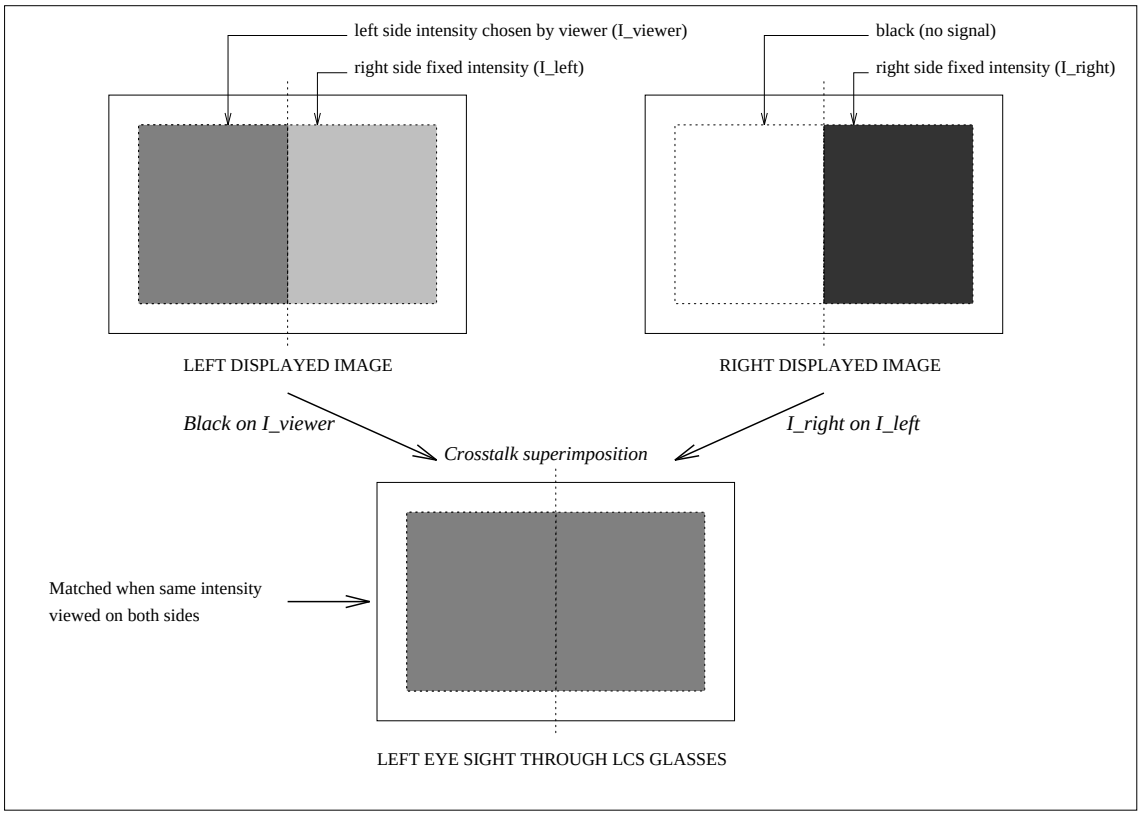


Figure 14: The experiment

and through the LCS glasses (we considered that the effect on the right eye was the same as for the left one). On the right part of the screen, two intensities for each component were chosen for each step of the experiment. First, the intensity of the left "wanted" image, and second the intensity of the right "unwanted" image inducing crosstalk. The viewer could only control the left part of the screen. He could change the intensity of the colour of the image for the left eye, superposed with a black right image (no crosstalk). The goal was to obtain the same visual impression for both parts.

On the left part of the screen, we have :

$$G_{left,viewer} + \varphi(G_{left}, 0_{right}) = G_{left,viewer} + 0 = G_{left,viewer} \tag{19}$$

since no intensity creates no crosstalk, that is :

$$\varphi(C_{direct}, 0_{crosstalk}) = 0.$$

On the right part of the screen, we have :

$$G_{left} + \varphi(G_{left}, G_{right}) = G'_{left}. \tag{20}$$

So, when the viewer has chosen the intensity of the left part to perceive the same intensity on each side of the screen, we have :

$$G_{left} + \varphi(G_{left}, G_{right}) = G'_{left} = G_{left,viewer}, \tag{21}$$

$$\varphi(G_{left}, G_{right}) = G_{left,viewer} - G_{left} \tag{22}$$

Thus, we were able to determine the crosstalk function φ.

	0	30	60	95	135	185	235
0	0	20	36	51	63	75	85
25	0	2	13	27	40	51	60
50	0	0	3	10	20	29	37
75	0	0	2	5	11	19	24
100	0	0	1	3	7	12	17
150	0	0	0	1	4	7	10
235	0	0	0	0	2	3	6

Table 1: Crosstalk measurement for the green component (for one experiment)

Three persons did the experiment. The results were quite the same for each of us, showing that the crosstalk effect depended only on the display system, and not on the viewer. 49 points per component and per person were determined. The test grid was first regular. 6 levels for the direct intensity and 6 ones for the indirect and thus crosstalk intensity. But with regard to the results, the test grid was modified to fit with the variation of the crosstalk function. The Table 1 is an example of a

characterisation. In the first column we have the direct left image intensities, and in the first line we have the right indirect (crosstalk) intensities. We can clearly see that the impact of a dark right image on a bright left one is very low. On the contrary, a dark right image will create an important crosstalk on a dark left image. Inthat latter case, crosstalk elimination will be practically impossible.

The acquired surfaces were approximated by a fourth order 2 dimensional polynome using least squares (with Matlab). The differences between the experimental points and the approximated surface were less than 2%. For any component, we have the surface equation :

$$\varphi(x,y) = ax^4+bx^3y+cx^2y^2+dxy^3+ey^4+fx^3+gx^2y+hxy^2+iy^3+jx^2+kxy+ly^2+mx+ny+o$$

with x for the direct intensity and y for the intensity inducing crosstalk. The coefficients for the red, green and blue components are given in Table 2.

	Red	Green	Blue
a	6.8734e-09	1.0635e-08	1.7513e-08
b	-9.2313e-09	4.1037e-09	8.7318e-10
c	-4.6767e-08	-6.1770e-08	-6.2243e-08
d	-6.3485e-08	-3.1338e-08	-9.9951e-09
e	1.2353e-07	6.8764e-08	4.8712e-08
f	-2.0822e-06	-7.1724e-06	-8.8738e-06
g	1.6836e-05	1.6478e-05	1.6533e-05
h	4.1093e-05	3.4561e-05	2.6284e-05
i	-6.1612e-05	-3.9198e-05	-3.0531e-05
j	-6.3560e-04	6.7490e-04	8.5609e-04
k	-0.0080	-0.0085	-0.0075
l	0.0097	0.0072	0.0060
m	0.4045	0.4416	0.3985
n	-0.5105	-0.4304	-0.3763
o	4.4977	4.4528	3.8673

Table 2: Coefficients for crosstalk approximation functions

We can see that the results of the experiments were quite alike for the green and blue components. As for the red one, the crosstalk was not as important. The explanation could be that the LCS glasses do not have the same response for each component, but we also have to consider the characteristics of the monitor, and maybe also some psychovisual parameters.

4.2.3 Eliminating Crosstalk

Given the crosstalk functions for each component, the aim was now to try and eliminate the defect. First, we must remember that the right image has an impact on

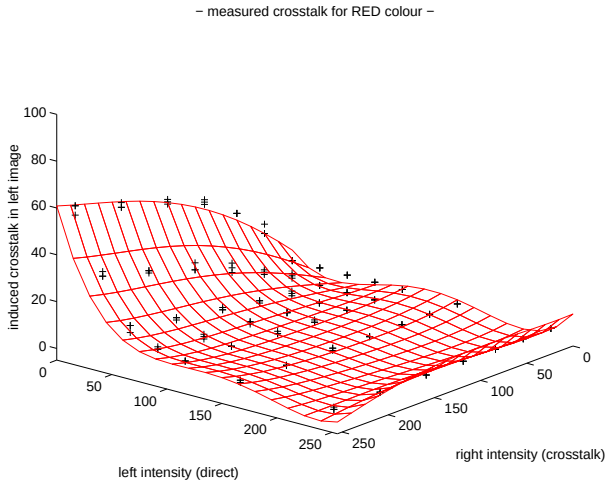


Figure 15: Measured crosstalk for red component

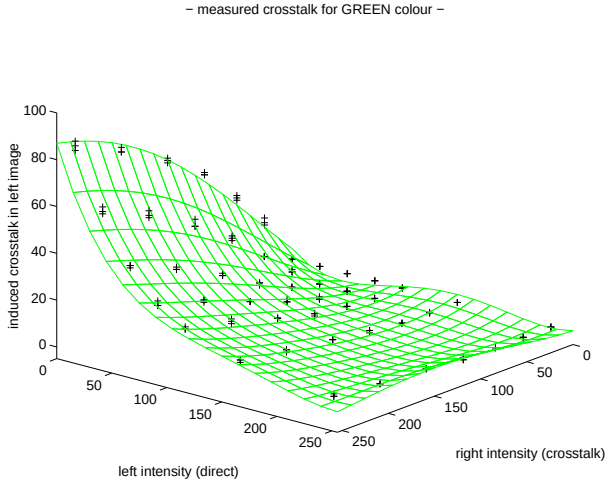


Figure 16: Measured crosstalk for green component

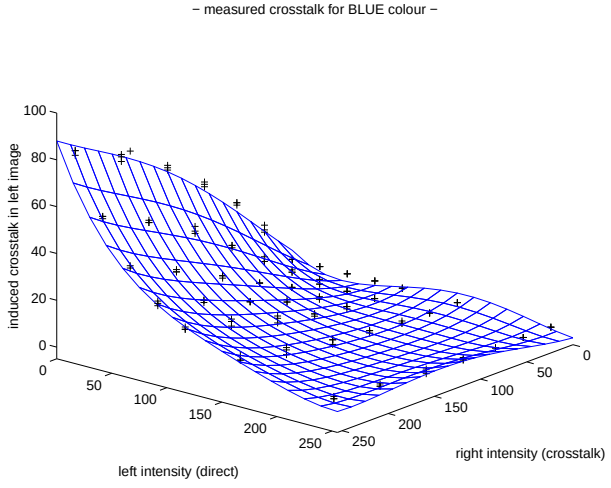


Figure 17: Measured crosstalk for blue component

the left one, but the reverse situation is also true. So crosstalk effect has to be eliminated considering both images at the same time. Without rectification, we have the equations (17) and (18). But what we now want to find is what intensity we must display for each left and right superposed pixel so that the left eye seems to see its left "monoscopic" image, and the right eye its "right" one, thus eliminating the ghosts and echos. So we must find $G_{left}^{displayed}$ and $G_{right}^{displayed}$ according to the equations coming from our model :

$$G_{left}^{displayed} + \varphi(G_{left}^{displayed}, G_{right}^{displayed}) = G_{left} \tag{23}$$

$$G_{right}^{displayed} + \varphi(G_{right}^{displayed}, G_{left}^{displayed}) = G_{right} \tag{24}$$

Clearly, to solve these equations, a non-linear method must be used. The method used is a recursive approach. We know that the displayed intensity are not that different from the expected ones on the screen. So, we consider :

$$G_{left}^{displayed}(n+1) = G_{left} - \varphi(G_{left}^{displayed}(n), G_{right}^{displayed}(n)) \tag{25}$$

$$G_{right}^{displayed}(n+1) = G_{right} - \varphi(G_{right}^{displayed}(n), G_{left}^{displayed}(n)) \tag{26}$$

with $G_{left}^{displayed}(0) = G_{left}$ and $G_{right}^{displayed}(0) = G_{right}$.

The convergence proved to be fast. For $n = 3$, we almost reached the convergence values (difference less than 2 units on 255). We were now able to apply our model on the stereoscopic sequences.

4.2.4 Results and comments

The principle of the elimination is in a way to subtract the unwanted crosstalk in the images, to avoid an over-intensity of the components in some areas. So if we want to eliminate the crosstalk, we must have the ability to diminish the intensity in the wanted image. Here is the main limitation of the rectification. If for a given pixel we have an important crosstalk and if the intensity of the pixel to be rectified is very low, we will not be able to eliminate the "echo" in the image at a hundred percent. For instance, if we have a very bright object on a dark background and if we have an important parallax for this object, it is obvious that the ghost will no totally disappear. So, if we have a component with an intensity at least 50 on 255, all ghosts for this component will practically disappear. The elimination is quite good, the left and right images become more accurate, and the three-dimensional view really increases in quality. But in the case of a component with and intensity less than 30 on 255, the crosstalk will remain if we have a very bright superposed image.

When rectifying stereoscopic sequences, two methods were used to prevent the recursive calculation from giving "negative" intensities and try and eliminate the crosstalk with the best results. The first one consisted in giving a minimum for the intensity (e.g. 20 on 255). Any pixel with a component intensity less than this

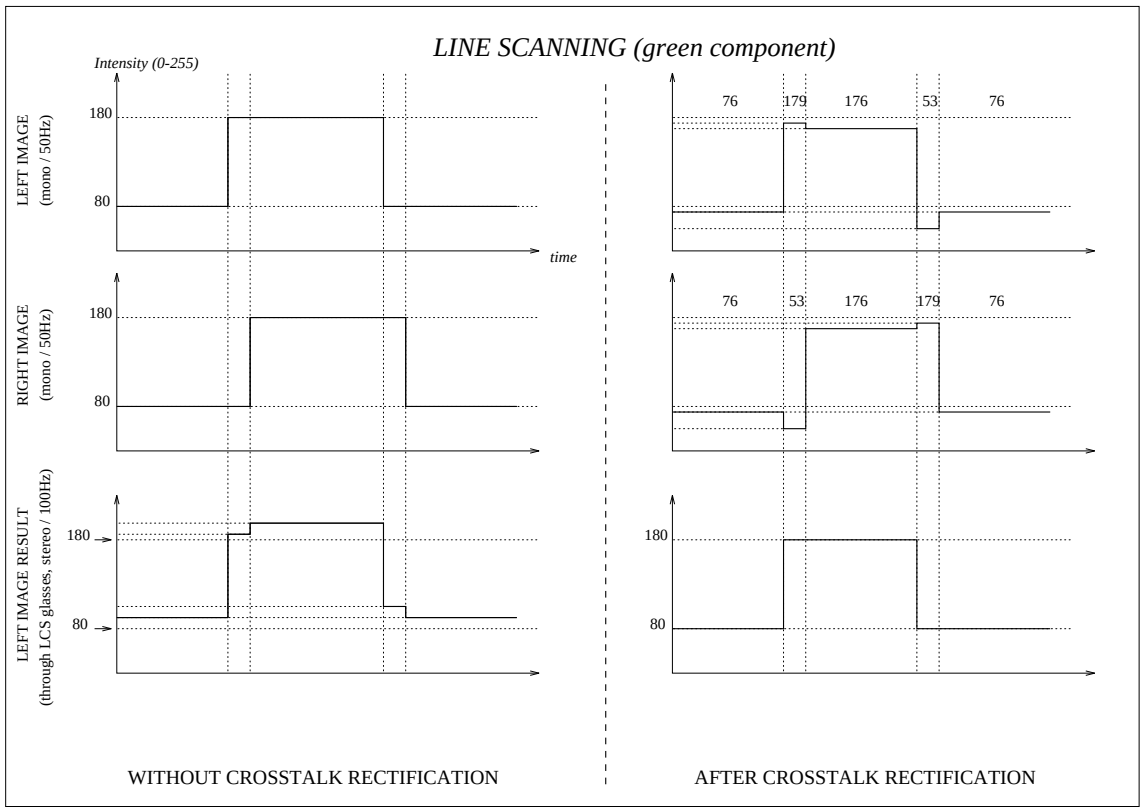
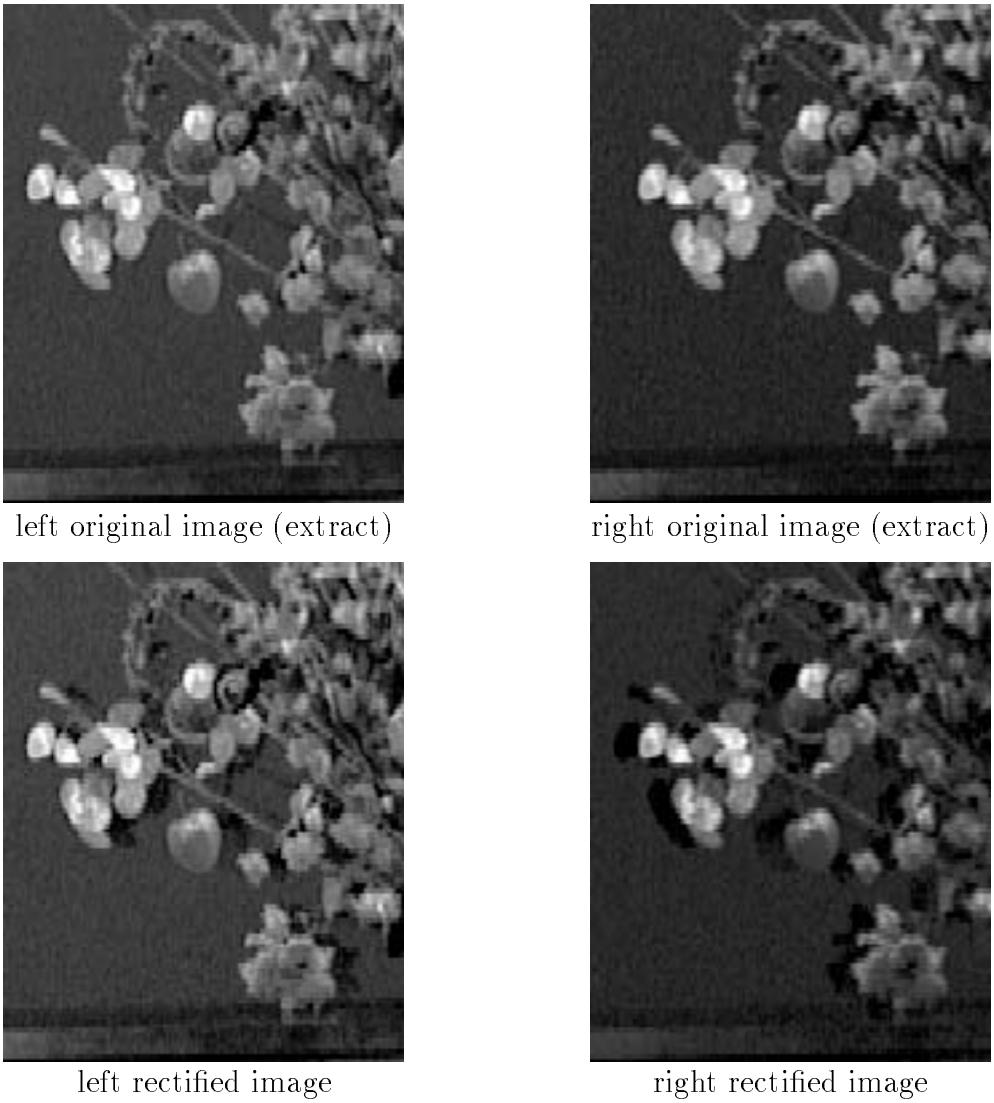


Figure 18: Example of efficient crosstalk elimination

minimum was given the value of the minimum. The second one consisted in remapping the $[0, 255]$ interval in the new interval $[20, 255]$. These modifications did not much change the aspect of the images and allowed a better rectification of the crosstalk, since they allowed a higher subtraction of the undesired defect.



left original image (extract) right original image (extract)

left rectified image right rectified image

Figure 19: An application of crosstalk elimination with good results

5 conclusion

If stereoscopic video is nowadays very attractive but not so much developped, that is because of its numerous unknown aspects required to use it. The grammar of 3D is not complete. The defects described in this report show that there is a lot more to do to improve the quality.

Concerning the keystone effect, there are two important things to notice. Firstly, the rectification would be useful for short convergence distance, to avoid vertical parallax which causes a discomfort for the eyes of the viewer. But we came to the conclusion that for the majority of the sequences, this is not of our concern. Secondly, eliminating vertical parallax means that we can only consider horizontal parallax as major difference between a left and right image. This can be very useful in the field of stereoscopic images compression, since the images could be studied only in 1 dimension, line by line. Thus we could have fast and high compression of the sequences, by taking mainly into account the horizontal disparity of the left and right images.

As for the crosstalk distortion, the results really depend on the characteristics of the images, but the model we studied seems to be a good step toward better viewing conditions for time-sequential display systems. Maybe the 3D home TV set is not that far away from us...

References

[1] O. Faugeras, Three dimensional computer vision: geometric viewpoint, 1991

[2] L. Lipton, CrystaEyes handbook, StereoGraphics Cooperation, 1991

[3] S. Pastoor, 3D-television: A survey of recent research results on sub jective requirements, Signal Processing: Image Communication 4,21 -32, 1991

[4] A.Woods, T.Docherty and R.Koch, Image distortions in Stereocopic System s, Proceeding of the SPIE Volume 1915, Stereoscopic Display and Application IV, February,1-13, 1993

[5] Robert G. Keys, Cubic Convolution Interpolation for Digital Image Processing

[6] Cheng Hong Yang, Geometric Models in Stereoscopic Video, Rapport technique de l'INRS-Télécommunications no. 95-12