STEREO IMAGING, AN EMERGING TECHNOLOGY

*E.A.Edirisinghe, **J.Jiang
*Department of Computer Science, Loughborough University, UK.
**School of Computing, Glamorgan University, UK.

Abstract

The presence of a third dimension, coupled with the advance functionalities of modern image / video-coding standards, could add a feeling of reality to the two dimensional images we have learnt to admire for decades. As a result of the rapid development of the internet, digital entertainment media and virtual reality, recent years have shown an increased interest in research and development related to stereoscopic imaging. The advances in LCD technology has already led to the design of several promising prototypes of free viewing stereoscopic displays. These have been an inspiration to many researchers working in the areas of transmission, storage and standardisation issues related to stereo imaging. In this paper we present the state of the art in stereo imaging technology, giving special emphasis to display and transmission technologies. Finally, the possible applications of this cutting edge technology to electronic business, science and education are discussed.

1. Introduction

One of the most amazing properties of the human vision system is its ability to feel the depth of the scenes being viewed. This is made possible by a process named *stereopsis.* It is the ability of our brain to fuse together the two images seen by the eyes (the *stereo image pair*) to form a single image, named the *cyclopean image* that contains embedded information about depth and an improved resolution of detail.

Under normal viewing conditions, the world appears to us as seen from a *virtual eye* placed midway between the left and right eye positions. This simple geometrical arrangement has an important consequence: since the world appears different from any of these viewpoints, the image we perceive of the world is never recorded directly by any sensory array, but constructed by our neural hardware. However, it is possible to stimulate our sense of stereo vision artificially by acquiring two pictures of the same scene and then presenting the left image to the left eye and right image to the right eye, allowing the brain to fuse them together for three-dimensional (3D) perception. The art and science of capturing, storing, editing, transmitting and displaying such 'still' and 'moving' images is defined as *stereo imaging*.

The earliest interests in stereo imaging were directed on stereo photography. The invention of the first binocular camera by Sir David Brewster in 1949, resulted in a huge trade in stereoscopes and stereo images. The start of London Stereoscopic and Photographic Company in 1850 and the formation of the Stereoscopic Society in 1893 are considered two early milestones in stereo imaging. Subsequently, with the 3D movie craze of the early 1950's and the wonder of holography in the 1960's, the excitement about the feeling of depth and reality grew exponentially, leading research teams around the world to explore the possibilities of making free viewing, "ultimate 3D experience" systems. After the rapid development of the computer graphics industry in the 1990's, there was a general realisation that the two-dimensional projections of three-dimensional scenes, traditionally referred to as "three-dimensional computer graphics", are insufficient for inspection, navigation, and comprehension of some types of multivariate data. For such data, the often neglected human depth cues of stereopsis, motion parallax and to a lesser extent, ocular accommodation, are essential for an image understanding.

The increasing demand in consumer products has resulted in the identification of number of new application areas where such additional information can be made use of. The traditional driving force behind the development of stereo imaging technologies, the entertainment industry, still remains a major driving force. However, the markets for 3D games, VR simulations, 3D graphic interfaces, the efforts for the infrastructure development of specifying data structures for transmitting 3D objects (VRML) and 3D video conferencing (RACE and ACTS) strive to drive the technology into greater heights. The recent advances of liquid crystal (LC) technology has finally resulted in the design of sophisticated free viewing systems, that would soon be developed in to ultimate real 3D display products [27]. Industrial research organisations such as NHK, HHI, DERA, Phillips, Sanyo, Sharp, Xenotech, Reality Vision, DTI, Nu Vision and academic research groups at MIT, Cambridge, De Montfort, Curtin and Tsukuba Universities are carrying out research in developing future stereo image display technologies. Many other research groups around the world are developing supporting technologies for the capture, storage, editing and transmission technologies related to stereo imaging. The new programmable multimedia signal processors, the introduction of high speed digital networks and the ever increasing computing power efficiencies, would lead this technology in the years ahead.

In this paper we provide the results of an in depth survey of stereo imaging technologies. Section-1, is an introduction to the subject area of stereo imaging with a brief historical overview. Section-2 provides a detailed description of stereoscopic image capture technologies. Section-3 provides a detailed analysis of various stereoscopic display techniques that have been introduced and developed over the years. Section-4 introduces the reader to various data reduction techniques that can be employed in association with the above capture and display technologies. Section-5 identifies the human factor considerations related to stereoscopic imaging technologies. Section-6 provides an insight into the application areas of stereoscopic imaging and possible future developments. Section-7 concludes, summarising the issues covered in this paper.

2. Stereoscopic Image / Sequence Capture

In Section-1 it was mentioned that stereovision could be simulated by acquiring two views of a 3D scene and by presenting them separately to the left and right eyes. Stereo capture equipment play an important role in the first step of this process. Many technical and operational variations are present between the practical designs of capture equipment. In the following sections we introduce the reader to the basic principles of some capture technologies that have been patented mostly in Japan, US and Europe [1].

2.1 Single Camera Capture

An obvious way of capturing a stereo image pair is to take a picture of the same scene from two horizontally separated points using the same camera. Recently, the application of precise optical, electro-optic, mechanical and electronic means to improve camera calibration has resulted in the application of this basic technique in robotics and stereo photogrammetry. The frequency, polarisation and spatial statistics of one or more light beams projected upon an object have also been analysed by multiple elements of a single sensor camera to obtain positional information. Alternatively, two images could be obtained from the same camera, under different polarisation and complete overlap, and subsequently separated. For close-up stereo imaging, such as stereo-endoscopy, mechanical and electro-optical devices that divide the camera lens into left and right halves have been used. A more frequently used technique has been to place an optical adapter in front of the existing lens so as to separately capture the two views. One distinctively different approach to the above mentioned techniques is to use camera-lens combinations that cyclically change their refractive index to capture the scene at varying depths. The latest single camera stereo image capture techniques involve the extraction of depth information from a single point of view combined with other information about the scene.

2.2 Dual Camera Capture

An important condition in capturing a stereo scene with two cameras is to align them within close tolerances in all three axes. Furthermore, the zoom, focus and the inter-axial distance or the convergence point of the two optical axes needs to be precisely controlled. Early work in this area consists of mechanisms for manual interlock of focus and convergence, or of zoom and focus, or of manual adjustment of one parameter at a time for both cameras. Naturally, the more recent attempts have been to automate these functions with application specific circuits or with programs written into a dedicated microprocessor unit. Image processing and storage techniques have also been used to reduce excessive horizontal parallax, eliminate camera shake and misalignment and to compensate for binocular asymmetries that result from zooming. Many schemes have been introduced for camera switching, digital storage and/or processing or novel display techniques to improve the actual or apparent vertical resolution. Mixing arbitrary numbers of left and right eye fields, using two cameras with electro-optic shutters and a single common optical element to facilitate synchronous zooming, are two such techniques. One or more high resolution black and white cameras have been combined with a low resolution colour camera to give a high resolution stereo image that would be otherwise unobtainable or expensive to produce. An alternative design has used an ultrasonic sensor to adjust the camera parallax to a viewers position. Stereo cameras that automatically track objects and adjust the zoom to keep them centred are being used in military, advance security and surveillance applications.

However, for stereoscopic video applications the most commonly used capture technique is the use of a pair of *genlocked* cameras in association with a stereo multiplexer that provides field sequential stereo for recording and viewing stereo with any CRT. A stereo demultiplexer could also be used to separate out the left and right images for dual video projection viewed with passive polarised glasses. Technologies have also been developed for stereo data to be recorded on one tape, in side by side or above/below compressed format [2].

3. Display Technology

The ultimate aim of a stereoscopic imaging system is human perception, in three dimensions. However, many variations are present between the methods in which the left and right views can be directed to the corresponding eye. We classify stereoscopic display units into three groups, namely, *Off-Head Displays* (OHDs), *Head Mounted Displays* (HMDs) and *Autostereoscopic Displays*. The OHD's and HMD's require optical devices close to the observer's eyes, whereas, the autostereoscopic displays have the eye-addressing techniques completely integrated into the display itself.

3.1 Off-Head Displays

The OHDs currently in use usually require that the devices for the image separation be worn by the observer. These devices are used to direct the left and right optical signals to the appropriate eye by the use of three multiplexing techniques.

Colour multiplexing (Anaglyph): This is one of the primitive kinds of viewing aids used for observing stereoscopic photographs. The left and right-eye images are filtered with near-complementary colours (e.g. red / green) and the observer wears respective colour-filter glasses for separation. However, colour rivalry effects and visually unpleasant transitory shifts in chromatic adaptation, restrict the use of this method.

Polarisation glasses: Polarisation glasses (linear or circular) in combination with orthogonally polarised images presented on two monitors provide a solution to the colour rivalry effects of the anaglyph technique. However, over 60% of the light emitted would be lost through the filters, and the remaining light flux is almost halved by beam combiners that have to be used. However, polarisation techniques are still widely used in viewing stereoscopic video projection screens. Recently LCD-based, direct-view displays and overhead panels have been commercially made available [Phillips].

Time multiplexed displays: These displays make use of the capability of the human visual system to merge the constituents of a stereo pair across a time-lag of up to 50 ms [3]. The left and right-eye views are shown in rapid alternation and synchronised with a LC shutter, which opens in turns for one eye, while occluding the other eye. The shutter-system is controlled via an infrared link. Both constituent images are reproduced at full spatial resolution by a single monitor, thus avoiding geometrical and colour differences.

Time sequentially controlled polarisation: This is essentially a combination of time and polarisation-multiplex techniques [4]. The idea is to switch the polarization of a LC panel on the monitor in synchronism with the change of the left and right-eye views. This approach using inexpensive and light-weight polarizing glasses offers two useful advantages over systems with active shutter-glasses. The LC panel can be constructed from several segments which operate independently on the active portions of the screen, hence ensuring that each eye is only reached by the intended image contents, thus reducing cross talk. Multiple display arrays can be operated without any extra synchronizing circuitry.

3.2 Head Mounted Displays

HMDs which are commonly sighted in virtual reality and 3D visualisation, use location multiplexing, in which the two views are created at separate places and relayed to the appropriate eye through separate channels. The image plane appears either in a fixed accommodation distance or in a variable, gaze-controlled position. Usually, the natural surroundings are optionally occluded by sight, giving rise to a feeling of total immersion in the displayed scene. Head tracking devices are used to create eye-point dependent changes in perspective when the user moves, thus preventing lost of view under head movements. This avoids musculo-skeletal problems. On the other hand, latencies and tracking-errors tend to provoke odd, uneasy sensations through conflicting visual stimulation and postural feedback, and adaptation can lead to reciprocal after effects. To release the user from the encumbrance of wearing an HMD a device named as Binocular Omni-Orientation Monitor (BOOM) has been designed [5]. Here the observer regards the stereo images as if looking through a pair of binoculars. A recent development in HMD technology is the "3-D display with accommodative compensation" (3DDAC) [ATR Labs]. Here the convergence distance is sensed by an infra red gaze-tracker and the position of a set of movable relay lenses is constantly adjusted, so that the screen surface (image plane) appears in the same distance as the convergence point of the gaze-lines.

The HMDs generally offer larger fields but suffer from weight concerns, resolution problems, computing power requirements and user discomfort (headaches and nausea). Recently, a number of free-viewing 3-D displays (autostereoscopic) have been designed with the idea of solving most of these drawbacks. The following section provides a detailed description of such techniques which may well hold the key to the widespread use of stereo imaging technology in the near future.

3.3 Autostereoscopic Displays

These type of displays can be broadly categorised into three groups, namely, electro-holographic displays, volumetric displays and direction-multiplexed displays.

Holographic techniques: These have the ability to store and reproduce the properties of light waves. The accuracy at which this can be done is very high, making it ideal for the production of free viewing 3-D displays. Attempts have been made to use acoustic waves to create a moving hologram in acousto-optic material [MIT]. Further, moving holograms have been created using optically addressed spatial light modulators [DERA Malvern]. However, coherent light is needed for illuminating the scene and the recorded interface pattern during recording and reproduction. This is a major shortcoming of this technique that limits its use to applications where the stereo scene is available only under special viewing conditions [6].

Several research institutes are pursuing research in adapting the principle of holography to a LCD environment [7]. However, the extremely high spatial resolution requirements of the LC panels that could be used in this regard and the handling of enormous amounts of data contained in a hologram are still open problems that need to be solved.

Volumetric displays: These can be divided into two fundamental groups, depending on where the image points representing a 3D scene appear to be seen. With the first type of system, a self-luminous or light reflecting medium (disc, spiral, helix shaped screens) is used, which either occupies the volume permanently or sweeps it out periodically [8]. The second type of volumetric displays is the multiplanar display type that creates aerial images in free space. The images representing various depth levels are written time-sequentially to a stationary CRT, whereas the observer looks at the screen via a spherical mirror with varying focal lengths [9]. In volumetric displays the portrayed objects appear transparent, since the light energy addressed to points in space cannot be absorbed by foreground pixels. Thus, practical applications seem to be limited to fields where the objects of interest are easily iconized or represented by wire frame models.

Direction-multiplexed systems: These can be broadly classified into four groups depending on the optical effect used for directing the rays emitted by pixels of the two perspective views. In the following sections we describe the principles of these techniques.

Diffraction based direction multiplexing:

Diffractive Optical Elements (DOE) approach – The corresponding elements of adjacent perspective views are grouped in arrays of so called 'partial pixels'. [10]. A diffractive grating placed in front of each of these arrays direct the incident light into the respective perspective images, viewing area. Advanced displays of this type integrate the above two stages within a single, high resolution spatial light modulator [11].

Holographic Optical Elements (HOE) approach – Holographic methods are used to model the optical properties of conventional optical elements. In a recent design [7], light from an ordinary light source contained in a backlighting unit is directed towards each eye by a holographic pattern recorded on a plate fixed to the back of a LCD. The unit ensures that the light passing through even-numbered rows is directed towards the left eye, and vice versa. Thus, provided that the viewer's eyes are in the correct area, they will see separate images.

Refraction based directional multiplexing:

Integral imaging - The scene is captured by an array of small convex micro lenses [12]. Each micro lens captures the scene from a slightly different perspective giving rise to an array of small 2D images of the same scene. An array of micro lenses of the same kind is used at the display. The light from each image point is emitted into the viewing zone as a beam of parallel rays at a specific direction by positioning the image plane at the focal plane of the micro lens array. Thus, the observer perceives different compositions of image points at different points of view. Since each pixel is spread to the lens diameter at replay, and the image formed behind each micro lens should be as complete and detailed as possible, the micro lenses must be as small as possible.

Lenticular imaging – The scene is captured by a set of vertically oriented convex lenses that are formed and placed adjacent to each other in a vertical plane [13]. The light from each image point is emitted at a specific direction in the horizontal plane. In the vertical plane they are emitted non-selectively, thus, changes of perspective in accordance with vertical head movements cannot be achieved by optical means. However, head tracking and computational image processing is used to reduce these effects.

A significant amount of research has been carried out with regards to this form of stereo image displays. Such displays can be broadly divided into two groups namely, direct-view lenticular imaging and rear projection lenticular imaging [13]. In the first type a cathode ray tube (CRT) or a LC panel is used behind every lens to produce a vertically oriented left and right image pair, which is projected to the lenticular screen. Head tracking is done by moving the lenticular screen itself or my introducing a movable plane with vertical slits, in between the screen and the CRT. In the second type, the LC panel is replaced by the stereo projector [14]. A dual lenticular screen is used, the inner screen of which focuses the projected images onto vertically oriented left right image pairs which are formed on a translucent diffuser plate, placed between the two screens. The lenses of the front lenticular screen map the images into the appropriate viewing zones. Head tracking is done by moving the front lenticular frame or the projection unit. For multiple user viewing, more than one projection unit has to be used.

Field lens displays – The basic idea is to use one or two field lenses to direct the light rays from the scene to the left and right eyes, keeping the left view dark to the right and vice versa. Several implementations have been proposed by various research groups. In one such method, the images are first formed on two LCD panels placed at right angles to each other [15]. Two light sources illuminate these panels with the use of closely placed lenses. The resulting light rays are superimposed and directed to the right and left eyes by a half mirror (beam combiner) that is placed at the bisection of the angle between the LCD panels. An alternative proposal is to use two lenses separated by a prism mask in association with a single source of light [Dresden University]. The prism mask is specially designed so that it deflects alternating columns of light rays into the left and right viewing zones. In a third kind of field lens based display, a stereo projector is used to project the left and right images on to a field lens placed between the observer and the stereo projector [HHI]. Thus, the stereo images are not projected onto a physical medium, rather they appear as floating aerial images in front or behind the field lens. The position of the floating image plane can be changed by adjusting the focussing characteristics of the stereo projection optics. A gaze tracker is used to sense the observer's momentary point of fixation. The aerial image plane is moved to this position by corresponding focus adjustments. As the observer accommodates on the distance of the aerial-image plane, accommodation distance and convergence distance coincides like in natural vision.

<u>Reflection based directional multiplexing</u>:

Direction multiplexing is done by a special screen named as a Retro-reflective screen that reflects the rays of light only in their original direction [16]. A stereo image projector with lateral movement flexibility projects light through a normal mirror system to this screen, which separates and projects the tow images through a half mirror to the observer. System has the ability to locate the current head position of the observer and to make later adjustments to the projector and to change the angle of the half mirror.

<u>Occlusion based directional multiplexing</u>:

The basic principle is to use parallax effects to hide parts of the image from one eye but to make those parts visible to the other. In the following sections we provide three techniques that work on different occlusion principles.

Barrier-grid displays – A latest design of this type consists of two barrier grids [17-Sanyo]. One placed in between the LCD display and its backlight case. The second is placed in front of the LCD display on the observers side. The two barrier grids are set and moved in such a way to direct the left and right views to the two eyes. In some designs, barrier grids are formed using LCD panels making software based control and gaze tracking possible [18-NHK].

Parallax-illumination displays – The parallax effect is created by a lattice of very thin vertical lines of light placed at a distance behind a LCD panel [19-Dimension Technologies]. The parallax causes each eye to view only light passing through alternate image columns. The lattice is made with the use of a lenticular lens sheet. It focuses light from a small number of fluorescent light sources to a large number of illuminated lines of a translucent diffuser. Head tracking is performed by the use of multiple sets of light sources and incorporating a large field lens to the display. The position of the lights sources is changed by switching between sets of laterally displaced lights. For multiple viewing, multiple sets of light sources are operated simultaneously, each producing the image pair for a single observer. A recent design incorporates two sets of blinking lights that produces two laterally displaced blinking lines behind each column of the LCD [20]. This provides full spatial resolution for both images. An even more advanced type of design consists of multiple sets of flashing light lines that create a stationary multiview display with look around capability.

Moving-slit displays – A fast switching LCD screen placed in front of a CRT is used to produce a single slit that could be moved to scan the screen at the rate of about 60 Hz. It channels different perspective views to adjacent zones. Several designs based on the above principle have been proposed. One such technique provides viewing capability for 16 users [21]. It uses three large spherical lenses to direct complete images to these locations. However, a CRT that works at 960 Hz frame rate is required for this design. An alternative design uses separate electron guns with magnetic beam control to produce the moving slit effect [22].

4. Stereoscopic Compression Methods

Most of the stereo image capture and display techniques described above require the transmission of two perspective views between the capture and display equipment. Certain applications may also need to store the information captured by a stereo camera pair. However, due the redundancies present between the two perspective images, they can be compressed, thus making saving on transmission bandwidth and storage space requirements. In this section we introduce the reader to various stereo image and sequence compression techniques that have been proposed by various researchers world-wide [24].

4.1 Stereo Image Compression

The first proposed stereo image compression algorithm [Perkins] was to code the sum and difference of the two images. However, the performance of this technique decreases with increased disparity values. A modification to this method is to shift one of the images horizontally to the point where the cross correlation between the images of stereo pair reaches its maximum value [Yamaguchi]. The shifted image is then subtracted from its partner image and the difference is encoded. This method, assumes that objects in the scene have similar disparity values and thus is not particularly efficient. Another approach is to translate the row blocks instead of the whole image.

A*symmetrical* approaches that exploit the *suppression theory* by subsampling one of the two images has also been proposed [Perkins and Dinstein et al.]. Even though these methods are attractive they are not appropriate for applications requiring fine details perception. *Disparity compensation* represents another example of asymmetrical coding. Lukacs was the first to introduce disparity compensated prediction. The aim of disparity compensation is to estimate the disparities between the objects in a stereo pair, and use these estimates to remove the stereo redundancy. The disparity is usually calculated via a block-based scenario. First, the left image is independently coded. Then the right image is divided into non-overlapping blocks. Either fixed or variable size blocks are used. Each of these blocks is shifted horizontally and compared to the corresponding blocks in the coded left image using some (MSE, SAD, etc.) measure to determine the similarity between the two blocks. The most similar block, is the *disparity compensated prediction*, and the corresponding translation is the *disparity* for the block.

Given the right image that is to be encoded and a disparity vector field, there are a variety of coding strategies that may be used. For example, the residual of the disparity compensated prediction may be encoded and transmitted. This method is referred to as Disparity Compensated Residual Coding. [Yamaguchi et al.]. In an alternative approach, each block is either encoded using disparity compensated prediction, or the block is independently coded using an adaptive Discrete Cosine Transform (DCT) [Dinstein et al.]. The approach used for a given block is determined by the accuracy of the disparity compensated prediction. An alternative technique is to estimate the transform of each right image block, from the transform (DCT) of the matching left image block, using a linear function of the form y=Ax +B [Perkins].

Disparity compensation can be implemented using both fixed block size and variable block sizes. Fixed block size schemes have the unique advantage of eliminating the overhead that is required to specify the block locations. However, if multiple objects or an occluded object exist in a block, these schemes cannot fully exploit the stereo redundancy. One solution is to decrease the block size. However, this will increase the overall bit rate for the disparity field. As a solution to this, we proposed a pioneering block based predictive coding algorithm [25] that completely eliminates the necessity of coding the disparity field. Another approach is to use a segmentation approach. However, not all segmentation techniques are suitable due to the excessive number of bits required to describe the shape and the location of each region. Quad-tree decomposition provides a relatively economical and effective solution to this problem. The idea is to decrease the block size adaptively when the prediction fails [Seferidis et al., Aydinoglu et al., Sethuraman et al.]. As block matching cannot handle non-linear deformations such as perspective distortion, a generalised block matching algorithm that approximates the deformations of the objects, by deforming the corresponding blocks in the picture, has been proposed [Seferidis et al.]. A windowed disparity compensation scheme has been proposed to solve the presence of discontinuities between disparity compensated blocks in the above algorithms [Aydinoglu at al.].

Other geometric relations than the ones concerning disparity can also be exploited. For instance, information about *projective transformations* to reconstruct one viewpoint from the other can also be transmitted [Oisel et al.]. *Symmetrical approaches* that involve the creation of an intermediate representation can also be considered [Vleeschauwer]. A stereo image pair can be coded using Visual Pattern Image Coding (VPIC) due to the capability of its coding primitives to reflect meaningful physical properties of projected real-scene surfaces: high-information edge regions and uniform regions [Craievich]. In an alternative approach wavelet transforms have been used to obtain a multi-resolution analysis of the stereo pair [Reynolds Jr.]. The resulting sub bands convey the necessary frequency domain information, enabling the suppression theory to be used in the compression. Methods of improving the existing stereo image compression algorithms by incorporating vergence and spatially varying sensing have also been proposed [Basu et al.]. For their micro lens array based full parallax stereo image display (see section 3.0), Forman et al, have proposed the use of a 3D DCT based hybrid DPCM/DCT coding scheme.

Several region based algorithms have also been proposed. In one such method, two segmentation trees are constructed progressively in top-down order. Regions are validated if a satisfying match is found, and then their division is stopped in the trees [Randriamasy et al.]. An algorithm for extracting regions and subsequently labelling line segments (boundary, join or isolated) in images and then using the scheme for stereo matching has also been proposed [Huynh et al.]. In an alternative approach three types of regions are identified: occlusion, edge and smooth regions. The region in the right image that is occluded due to finite viewing area is independently coded. The non-occluded region is segmented into edge and smooth regions and disparities are estimated using a block-based approach. Subsequently a Subspace Projection Technique (SPT) is used as a post processing technique to remove the effects of photometric variations [Aydinoglu et al.]. In [26] we proposed an object-driven block-based algorithm for the compression of stereo image pairs. A contour based technique was used to identify matching objects in the image pair. Different coding schemes were applied to the boundary and internal areas of these objects and to the background area, to selectively code these areas at different quality levels.

4.2 Stereo Sequence Compression

For the case of stereoscopic sequences, many redundancy sources are simultaneously available, including intra-image structure, motion, and stereoscopy. Various methods, inspired from the MPEG compression standard, have been proposed to exploit stereoscopic redundancy.

A straightforward MPEG-2 based design is as follows. The left channel is coded by means of a DCT with motion compensation. The right channel is coded by motion compensation, disparity compensation, or direct block encoding with a DCT, choosing for each block the estimate yielding the smallest reconstruction error [Kopernik, Chassaing].

In order to avoid artefacts associated with block-based approaches when high compression rates are applied, it is also possible to treat the image information in terms of two-dimensional objects [Ziegler]. One such procedure is to segment the images into uniform regions, corresponding to objects, using available stereoscopic and motion information. Moving objects are then transmitted with a set of parameters describing their colour, form, motion, and disparity. Stereoscopic and dynamic information can also be integrated into a common and compact description of the scene, consisting of a 3-D structure and 3-D motion parameters [Dugelay et al., Grammalidis et al.]. Alternatively, object modelling has been used to facilitate motion and disparity estimation by taking into account, for instance, the difference between the current image and the projection of the estimated 3D model [Falkenhagen]. The transmitted information can also be represented in terms of 3-D objects composed of planar surfaces, with motion described by rotation and a translation component [Dugelay et al.]. Alternatively, 3D surfaces can also be represented by depth information at zero crossings, which have been obtained from a contour detector [H.Morikawa]. Even though 3D scene modelling represents an attractive coding approach for stereoscopic sequences, the complexity of this modelling is often very important. It is also possible that this modelling be inadequate in some image regions because of occlusions, for instance.

A different approach to stereo sequence compression is, explicit extraction of depth from the stereo data. One advantage of this method is that frontal objects can easily be separated from the background [M.Waldowski]. It then becomes possible to treat these two entities differently, using hybrid approaches, combining 2D and 3D representations [Grammalidis, Kopernik].

Multi-resolution techniques are another popular approach for stereo sequence compression. One such algorithm exploits the high correlation in stereoscopic image sequences by straightforward computing of blockwise cross-correlation's and multi-resolution hierarchical matching using a wavelet-based compression method [Siegel et al.]. In a second method, one of the bit streams is independently coded along the lines of MPEG 2 standard, while the other stream is estimated at a lower resolution from this stream. A multi-resolution framework is adapted to facilitate such an estimation of motion and disparity vectors at different resolutions [Sethuraman et al.]. A scheme that leads to disparity discontinuity preserving, yet keeping smoother and more accurate disparity fields than in fixed block size based schemes has also been proposed [Sethuraman et al.]. The segmentation is achieved by performing a quad-tree decomposition, with the disparity compensated error as the splitting criterion. In an alternative scheme proposed by the same authors, each frame in one of the streams is segmented based on disparity. The corresponding segments in the other stream are encoded by reversing the disparity map obtained during the segmentation. Areas without correspondence in this stream, arising from binocular occlusions and disparity estimation errors, are filled in using a disparity map based predictive error concealment method.

A compression scheme for interlaced stereoscopic sequences has also been proposed, which differentiates between a region of fixation and a peripheral area, and thereby compacts the stereoscopic information into the spectral space of a monocular video channel [Labonte et al.]. Spectral compression is achieved by avoiding transmitting high frequency information over entire images, but only within and around the region where the observer acuity is the highest. The output from a stereo disparity estimation process using calibrated cameras gives absolute 3D surface co-ordinates from a single stereo pair [Papadimitriou et al.]. When combined with monocular motion cues, the true 3D motion parameters of moving objects can be accurately calculated. Further analysis enables segmentation of body elements according to motion while the 3D surface feature structure, although available from the start, can be integrated.

An object-based stereo image coding algorithm that relies on the modelling of the object structure using 3D wire-frame models, and motion estimation using globally rigid and locally deformable motion models has also been proposed [Malassiotis et al.]. The authors also proposed a joint estimation of motion and disparity vector fields from stereoscopic image sequences by modelling the local interaction processes using Markov Random Fields (MRF). Several schemes based on motion modelling has also been proposed [Grammalidis]. The 2D motion of each object observed in one of the two views is modelled using a 3D-motion model involving a translation and a rotation. The regions where the 3D model is applied are identified using a motion based split and merge technique. Furthermore, an extension of the 3D motion estimation method that uses a single 3D motion model to describe the apparent 2D motion in both channels has been proposed. These 3D motion estimation methods are then integrated in a stereoscopic inter-frame coding scheme. A hybrid coder, using block-based coding as a fallback mode in cases where 3D-motion estimation fails, has also been proposed. An object-based coding scheme that is in many ways similar to the above method has been proposed by Tzovaras et al. A multi-resolution block-based motion estimation approach is used for initialisation, while disparity estimation is performed using a pixel based hierarchical dynamic programming algorithm. A split and merge segmentation procedure based on both 3D motion and disparity is then used to determine regions with similar motion and depth parameters. This is combined with an efficient depth modelling method that offers full depth information at the decoder site. Motion and depth model parameters are then quantized and transmitted to the decoder along with the segmentation information. An object-based motion compensation scheme is then used to reconstruct the original image, based on the objects created by the segmentation approach.

5. Human Factors in Stereo Imaging

Human factors issues, play an important part in the development of any technology into a mature status. Here we provide a summary of issues that need to be considered in the design and development of stereo imaging technologies [NHK, ATR-HIP, CCETT, HHI].

Previous research has shown that even amongst the majority of the people with good stereovision capability, wide variations are present in depth perception. Thus, in the design and testing of stereo imaging technologies, these effects need to be considered. It has been shown that with the frequency and duration of use, the stereovision capability of a person may improve, making binocular fusion rapid and more comfortable, thus making the subjective testing of stereo imaging technologies an almost impossible task. The viewing distance, screen size, horizontal and vertical parallax, binocular asymmetries, non stop depth cues will all result in the determination of the overall viewing comfort of a stereo scene. As an example, the size distortion of stereo images become noticeable in instances where angular size of a displayed object, and its perceived distance, do not covary as in real world conditions. As a result, smaller displays may contain annoying miniaturisation effects. Research has indicated that display sizes of 34” and 50” provide no distortion effects for television like, motion picture display [3]. Unlike under natural viewing conditions, large binocular parallaxes (disparities), tend to produce eye strain. Research has shown that disparities up to 35 minarc, do not cause any noticeable discomfort, irrespective of the spatial resolution of the constituent 2D images [3]. Further, cross talk between the left and right images produces double contours, which is known as ghosting, which in turn would result in headaches. It has also been shown that cross talk increases with the increase of contrast and binocular parallax. The complete removal of cross talk is a practical impossibility. However, for multiview displays having a limited number of views, cross talk could be an advantage. This is due to the fact that it results in the reduction of image flipping, which is essentially the noticeable jump of the image from one perspective to the next. Substantially large number of perspective views would be required to make both flipping and cross talk imperceptible.

The success or failure of a stereo imaging system design would largely depend on the visual comfort it provides to the viewer for the long duration viewing of high quality stereo images. Thus, human factors issues will remain an important part in the design of modern and future stereo imaging technologies.

6. Applications & Future

The interest in stereo imaging spans across a wide range of application areas. Unlike two dimensional technologies that we admired for decades, which attempts to display depth and perspective in a flat, two-dimensional environment, stereo imaging technologies provide its users a more realistic visual perception. Due to this fact it has already found many applications in engineering, architectural, scientific, educational and military fields. Stereoscopic imaging systems have been designed for 3D computer graphics, 3D visualisation, remote control vehicles and tele-manipulators, simulators and training systems, molecular modelling, computational chemistry and CAD. Further, these systems have been used to design next-generation automobiles and air planes, perform gene-splicing, in air traffic control, to interpret images gathered from deep space, in microscopic investigations of biological components, to perform endoscopic surgery and even to guide the Pathfinder mission on Mars.

At present enormous amount of interest is shown in the use of stereo imaging in digital entertainment media such as, computer games, virtual reality experiences, digital movies and television. However, these applications have so far made wide use of the existing two dimensional display media (TV and VDU) operated in field sequential mode for left/right views, in association with polarised eyewear glasses to stimulate the 3D effect. Such technologies have attained their maturity and have resulted in a standardisation effort, of a field-sequential 3DTV standard, to be used in association with the existing NTSC and PAL video standards. With the advancements in the design of autostereoscopic displays (see Section-3.3), free viewing systems for 3DTV capture, transmission and storage would be introduced to the consumer electronics market. However, issues such as, multi view capabilities (wide viewing zones), long duration viewing comfort, contrast constraints, higher definition and quality, information redundancy removal of the large amount of resulting data, are yet problems that need to be addressed before widely accepted commercial products intervene the present two dimensional television market.

Due to the advancement and popularity of Internet based technologies such as E-commerce, E-education, E-health, stereoscopic imaging will inevitably find many new application areas in future. The presence of the third dimension would be of particular importance in E-commerce where users could view true three dimensional images of goods that are available for sale. Complex laboratory experiments and demonstrations would have more meaning, when presented with true depth information for human visualisation and understanding. The ability to present true 3D pictures with full depth information, would help both patients and doctors in tele-medicine and E-health applications.

7. Conclusions

The fast evolving field of stereoscopic imaging is deemed to be the next revolution in digital imaging and entertainment media. With the advancement of stereoscopic imaging technology and the based technologies needed for it, many more scientific, engineering and day-to-day tasks would be solved and presented using this technology.

In this paper we have provided a detailed survey of stereoscopic imaging technologies, its history, applications and future potential.

References

1. M.Starks, “Stereoscopic Imaging Technology”, 3DTV Corporation, http://www.3dmagic.com/.
2. L.Lipton, “Stereo-Vision Formats for Video and Computer Graphics”, Stereo Graphics Corporation, http://www.stereographics.com/
3. S.Pastoor, “Human Factors of 3D Imaging”, HHI Berlin, http://atwww.hhi.de/.
4. P.J.Bos, “Time Sequential Stereoscopic Displays: The contribution of phosphor persistence to the "ghost" image intensity” Proc. 1991 ITE Annual Convention, Tokyo, 1991, pp. 603-606.
5. Bolas, M T, Lorimer, E R, McDowall, I E, Mead, R X “Proliferation of counterbalanced, CRT-based stereoscopic displays for virtual environment viewing and control.” Proc. SPIE 2177, 1994, 325-334.
6. S.A.Benton, “The second generation of the MIT holographic video system.”, TAO First Int. Symp., 1993.
7. D.Trayner, E.Orr, “Autostereoscopic display using holographic optical elements.” Proc. SPIE, 1996
8. H.Yamada, K.Akiyama, K.Muraoka, Y.Yamaguchi, “The comparison of three kinds of screens for a volume scanning type 3D display.”, TAO First Int. Symp., 1993.
9. D.McAllister, “Stereo computer graphics and other true 3D technologies”, Princeton University Press, Princeton, 1993.
10. T.Toda, S.Takahashi, F.Iwata, “3D video system using Grating Image.”, Proc. SPIE 2406, 1995, pp. 191-198.

11. E.Schulze, “Synthesis of moving holographic stereograms with high-resolution spatial light modulators.”, Proc. SPIE 2406, 1995, pp. 124-126.
12. M.McCormick, “Integral 3D imaging for broadcast.” Proc. 2nd Int. Display Workshop , Hamamatsu, 1995, pp. 77-80.
13. R.Börner, “Autostereoscopic 3D-imaging by front and rear projection and on flat panel displays.”, Displays 1993, 14(1), pp. 39-46.
14. D.Takemori, K.Kanatani, S.Kishimoto, S.Yoshii, H.Kanayama, “3-D Display with large double lenticular lens screens.”, SID 95 Digest, 1995, pp. 55-58.
15. D.Ezra, G.J.Woodgate, B.A.Omar, N.S.Holliman, J.Harrold, L.S.Shapiro, “New autostereoscopic display system.”, Proc. SPIE, vol. 2409, 1995, pp. 31-40.
16. “The Xenotech 3-D auto stereoscopic display system.” Technical Information. Xenotech Australia, Suite 1,41 Walters Drive, Osborne Park, Western Australia 6017.
17. G.Hamagishi, M.Sakata, A.Yamashita, K.Mashitani, E.Nakayama, S.Kishimoto, K.Kanatani, “New stereoscopic LC displays without special glasses.”, Asia Display '95, 1995, pp. 791-794.
18. H.Isono, M.Yasuda, H.Sasazawa, “Autostereoscopic 3-D LCD display using LCD-generated parallax barrier.” Japan Display '92, 1992, pp. 303-306.
19. J.Eichenlaub, “An autostereoscopic display with high brightness and power efficiency.”, Proc. SPIE 2177, 1994, pp. 4-15.
20. J.Eichenlaub, D.Hollands, J.Hutchins, “A prototype flat panel hologram-like display that produces multiple perspective views at full resolution.”, Proc. SPIE 2409, 1995, pp. 102-112.
21. A.R.L.Travis, S.R.Lang, J.R.Moore, N.A.Dodgson, “Time-multiplexed three-dimensional video display.”, SID 95 Digest, 1995, pp. 851-854.
22. T.Hattori, “Electro-optical autostereoscopic displays using large cylindrical lenses.”, Proc. SPIE 1457, 1991, pp. 283-289.
23. S.Pastoor, M.Wopking, “3D Displays: A Review of Current Technologies”, http://atwww.hhi.de/~blick.
24. E.A.Edirisinghe, “Data Compression of Stereo Images & Video”, PhD Thesis, Loughborough University, UK, December 1999.
25. J.Jiang, E.A.Edirisinghe, H.Schroder, “Algorithm for compression of stereo image pairs”, IEE Electronic Letters, Vol. 33, No. 12, 1997, pp. 1034-1035.
26. E.A.Edirisinghe, J.Jiang, H.E.Bez “An Object-Driven Block-Based Algorithm for the Compression of Stereo Image Pairs”, Proc. of the SPIE, Vol. 3808, 199, pp. 281-292.
27. M.Halle, “Autostereoscopic Displays and Computer Graphics”, ACM SIGGRAPH issue on Art of Display Technology, http://www.media.mit.edu/~halazar/autostereo/autostereo.html.

Biographies

Eran Anusha Edirisinghe received his B.Sc.Eng. degree in Electronic & Telecommunication Engineering from the University of Moratuwa, Sri Lanka, in 1994 and his M.Sc. degree in Digital Communication Systems, from Loughborough University, UK, in 1996. He completed his Ph.D. degree at Loughborough University in 1999. In November 1999, he joined the Digital Imaging & Data Compression research group of Glamorgan University, UK, as a postdoctoral research fellow. He has been an active researcher in stereo image compression for the past 5 years. His research interests include stereo and multidimensional image compression, computer graphics, pattern recognition, MPEG-4 video object coding and virtual reality.

Jianmin Jiang received his B.Sc. from Shandong Mining Institute, China, in 1982, his M.Sc. by research from China University of Mining & Technology, in 1984, and his Ph.D. degree from University of Nottingham, UK, in 1993. In 1989, he joined Loughborough University, UK, as a visiting scholar and later moved to the University of Nottingham as a research fellow funded by SERC in UK. He joined Bolton Institute, UK as a lecturer in Electronics in 1992 and later joined Loughborough University as a Lecturer. In 1998 he was appointed as professor of Digital Imaging and Data Compression at Glamorgan University, UK, where he has established a strong research group funded by national and international bodies. His research interests include, neural networks, video coding, stereo image coding, digital image databases, internet technologies and parallel computing.