

## 2 Stereo vision, 3D video capture and scene representations

Your goals for this “Introduction” chapter are to learn about:

- Different 3D video representations.
- Stereoscopic video capture technologies.
- 3D Image Warping/Depth Image Based Rendering (DIBR).
- The difference between left and right views based stereoscopic video vs. colour plus depth 3D video.

### 2.1 Different 3D video representations

3D objects can be reconstructed from the captured real world images, which provide the user the impression of 3D video. The methods of reconstruction and capture of the image sequences are based on the requirements of the targeted application scenario. According to the classification of MPEG-3DAV (Motion Picture Expert Group-3D Audio Visual), three scene representations of 3D video have been identified, namely omni-directional (panoramic) video, interactive multiple-view video (free-viewpoint video) and stereo video [2]. Omni-directional video allows the user to look around a scene (e.g. IMAX-Dome). This is an extension of planar 2D image into a spherical or cylindrical image plane. Figure 2.1 shows some example omni-directional images generated with the Dodeca™ 1000 camera system and post-processed with corresponding Immersive Media technology [3].

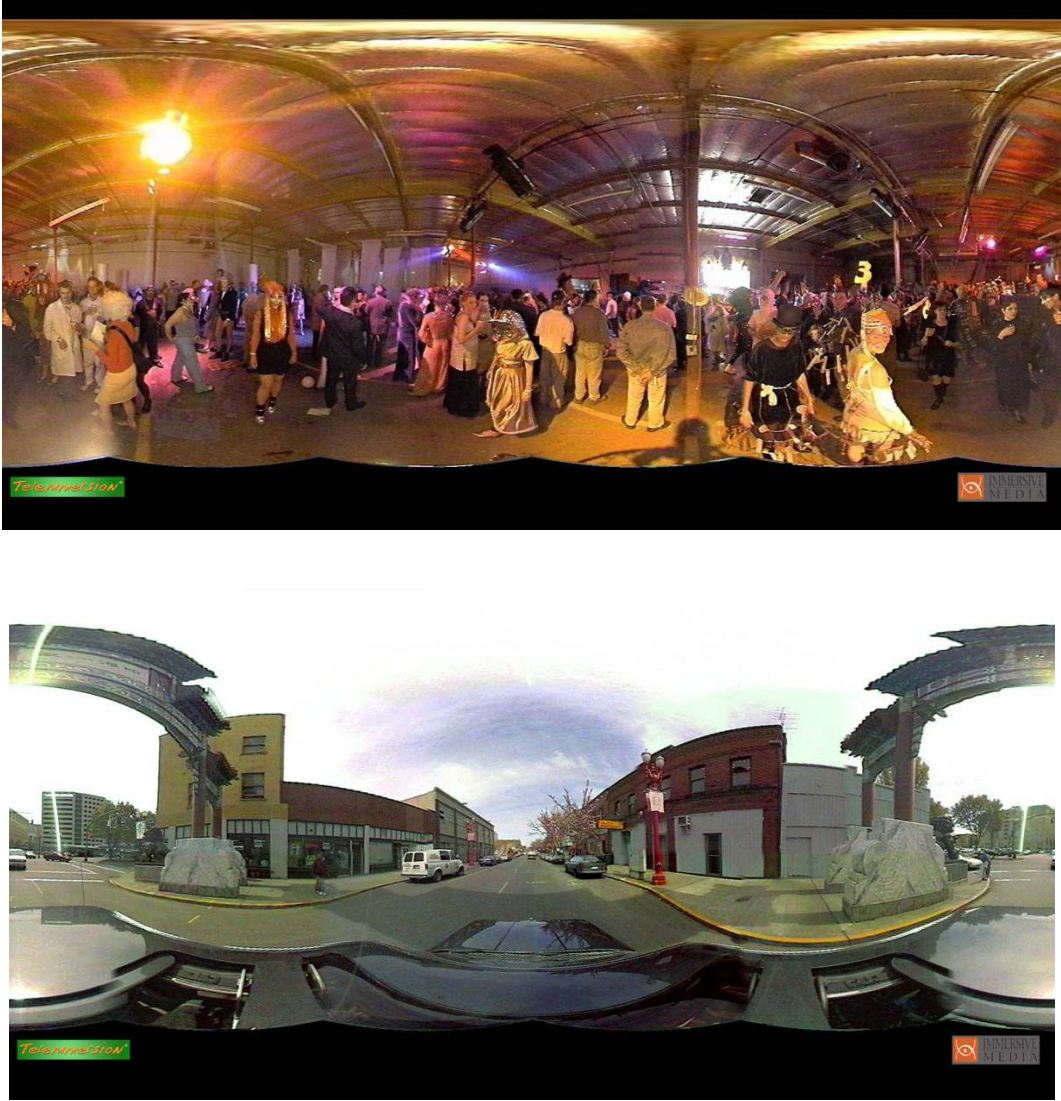



Figure 2.1: Omni-directional images from Telemmersion® video


The potential applications scenarios relevant to Omni-directional video are described in [4]. However, the 3D video in this format has a limited application scope (e.g. navigation and object avoidance) and may not be suitable for general application scenarios like 3D TV broadcasting. Multi-view video (e.g. free-viewpoint video) is the general case for all 3D scene representations. It allows the user to select an arbitrary viewpoint and direction within a visual scene, generating an immersive environment. It generates virtual camera views through interpolation of real camera views. This representation can be effectively utilized in wide range of applications, including FTV (free-viewpoint television) and surveillance [2]. The ray space approach and 3D model based approaches have been identified for real-time rendering of novel views [5]. Figure 2.2 shows an array of cameras (i.e.  $16 \times 16$ ) which can capture multiple raw video sequences and the captured multiple videos [6]. However, due to the high demand for system resources (e.g. processing power, bandwidth, and storage), the availability of multi-view video applications to the mass market will be further delayed till 3D video technologies and supporting infrastructure get to a more mature stage than the current stage of the development process. The third approach is stereoscopic video which we describe in more details in the next section.

SIMPLY CLEVER

ŠKODA

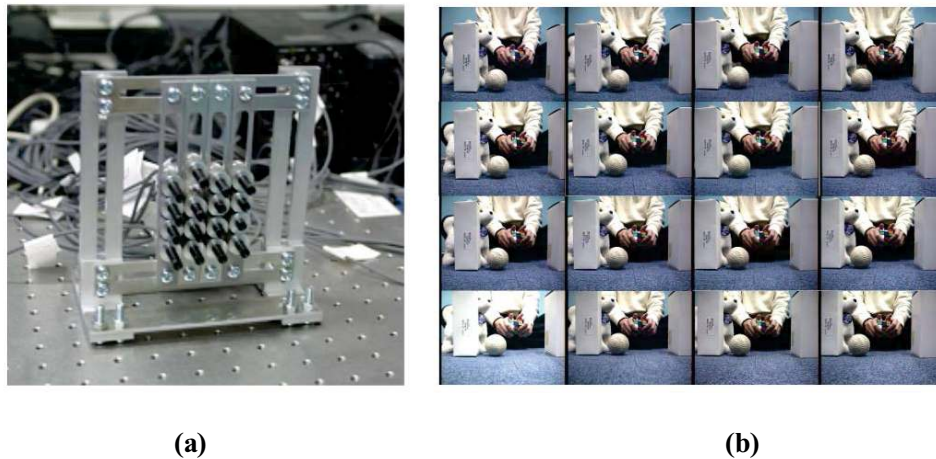


**We will turn your CV into an opportunity of a lifetime**



Do you like cars? Would you like to be a part of a successful brand? We will appreciate and reward both your enthusiasm and talent. Send us your CV. You will be surprised where it can take you.

Send us your CV on [www.employerforlife.com](http://www.employerforlife.com)



**Figure 2.2:** Multi-view camera platform; (a) Input system, (b) Captured images [6]

## 2.2 Stereoscopic video and capture technologies

In order to produce a 3D impression, stereoscopic video representation renders two adjacent views for each eye of the user. The left and right views are then fused in the visual cortex of the brain to perceive the depth of a scene (see Figure 2.3). According to [7], accurate perception of depth by human visual system can be attributed to two main categories of depth cues. The physiological and psychological depth cues mentioned in [7] are as follows:

### Physiological depth cues

- **Binocular disparity:** The dissimilarity in views due to the relative location of each eye.
- **Accommodation:** The change in the focal length of the lens in the eye caused by muscles in the eye to produce a focused image on the retina.
- **Convergence:** The rotation of eyes to align or merge the left and right eye images into a single image with perceived depth.
- **Motion parallax:** The difference in views produced by moving the scene or the viewer. For example, in a movie it is possible to realize the size of an object which is speeding towards the viewer based on the relative change in size with time. This cue often differentiates the realism of a video from that of a still image.
- **Chroma-stereopsis:** The difference in apparent depth due to the colour of an object from refraction effects in the eyes.



### Psychological depth cues

- Image size: This is a useful hint but not sufficient to determine size or depth of objects.
- Linear perspective: This is the decrease in the apparent size of an object with increasing distance.
- Aerial perspective: This refers to the hazy and bluish appearance of distant objects.
- Shading suggests that objects farther from the source of light are darker.
- Shadowing of an object on others provides clues about position and size.
- Occlusion: of objects provides a clue about their relative location.
- Texture gradient provides clues regarding distance and relative location.
- Brightness of an object suggests that it is closer than dimmer objects.

Stereoscopic video capturing system mainly exploits the binocular disparity cue which helps human visual system to perceive depth.



I joined MITAS because  
I wanted **real responsibility**

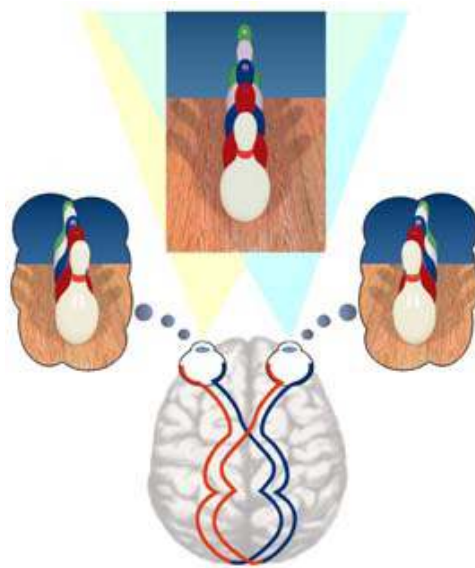
The Graduate Programme  
for Engineers and Geoscientists  
[www.discovermitas.com](http://www.discovermitas.com)

**Month 16**  
I was a construction  
supervisor in  
the North Sea  
advising and  
helping foremen  
solve problems

Real work  
International opportunities  
Three work placements

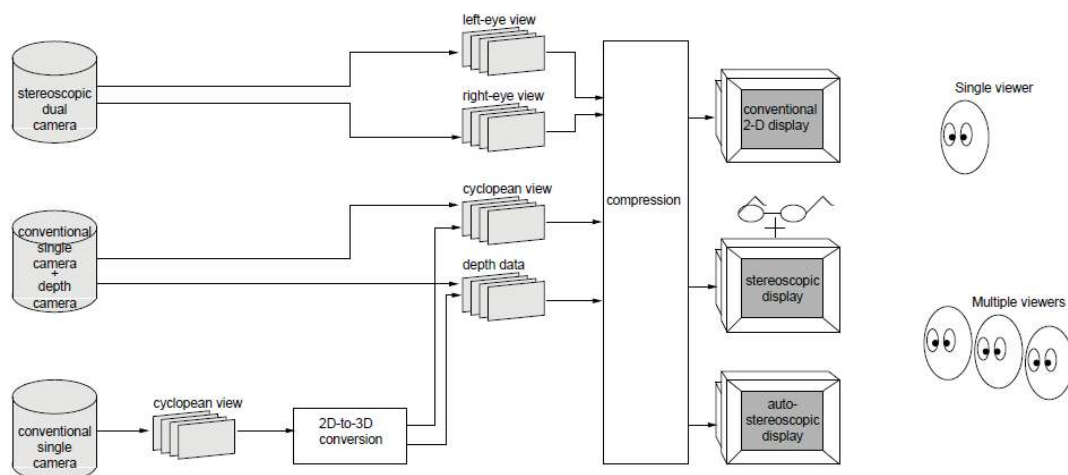


Stereoscopic video is one of the simplest forms of 3D video available in the literature. Moreover, this representation is a subset of multiple-view video, where only two camera views are rendered to the user. Due to the simple representation and adaptability (e.g. simple camera arrangement, cost effective display systems, etc), stereoscopic video could be employed in future broadcasting (e.g. 3D TV), storage and communication applications (e.g. 3D conferencing) relatively easily compared to other representations of 3D video. The existing infrastructure for audio-visual technologies (e.g. compression/decompression) can be adopted to send binocular content over communication channels. Moreover, the demand for resources (e.g. bandwidth and processing power) will be lower compared to the multi-view video. Therefore, in this book, stereoscopic video is considered as the main 3D video representation and the constraints and problems associated with stereoscopic video communications are presented.



**Figure 2.3:** Illustration of stereo vision [8]

At present, technologists are working on several 3D video applications, which cover the whole application chain consisting of 3D capture, compression, transmission, rendering of 3D video and high-end display technologies [9]. The separate modules in a stereoscopic video chain are shown in Figure 2.4. The term “cyclopean view” in Figure 2.4 refers to the intermediate view between the left and right view perspectives. In order to support stereoscopic video for single-user or multi-user display systems all the interconnected technologies (capture, transmission) should work in harmony. More importantly, the sufficient stereoscopic video content should be generated to meet the customer demand.



**Figure 2.4:** Separate modules in stereoscopic video chain [10]

There are several techniques to generate stereoscopic video material including dual camera configuration, 2D-to-3D conversion algorithms, 3D/Depth-range cameras [10]. Stereoscopic view of a scene captured using a stereo camera pair (i.e. the left-eye and the right-eye view are recorded separately by two cameras taken from a slightly different perspective) is the simplest and most cost effective way to capture stereo video at the moment compared to other technologies available in the literature (see Figure 2.5). The shooting parameters such as camera base distance (distance between the two cameras), convergence distance (distance of the cameras to the point where both optical axis intersect) and camera lens focal length can be utilized to scale the horizontal disparity and thus the degree of perceived depth. Furthermore, 3D video with the dual camera configuration provides fewer burdens at the receiver side for rendering 3D video due to the availability of two views. Two dual camera configurations can be distinguished, namely the parallel camera configuration and the toed-in camera configuration, also called converging cameras (see Figure 2.6). According to the study carried out in [11], the parallel camera configuration avoids geometrical distortions like the keystone distortion and depth plane curvature. As dual camera configuration generates two separate image sequences for left and right view, more system resources are necessary to process, store and transmit the generated content in comparison to the resource requirements of 2D video. For example, a double disk space is needed to store the raw left and right video sequences. Moreover, the viewing angle will be limited with the stereo camera pair and thus no interactivity can be employed. The commercially available stereoscopic video cameras and 3D add-ons for standard camcorders are listed in [12].



(a)

**ie** business school

#1 EUROPEAN BUSINESS SCHOOL  
FINANCIAL TIMES 2013

**#gobeyond**

**MASTER IN MANAGEMENT**

Because achieving your dreams is your greatest challenge. IE Business School's Master in Management taught in English, Spanish or bilingually, trains young high performance professionals at the beginning of their career through an innovative and stimulating program that will help them reach their full potential.

- Choose your area of specialization.
- Customize your master through the different options offered.
- Global Immersion Weeks in locations such as London, Silicon Valley or Shanghai.

*Because you change, we change with you.*

www.ie.edu/master-management | mim.admissions@ie.edu | f t in YouTube

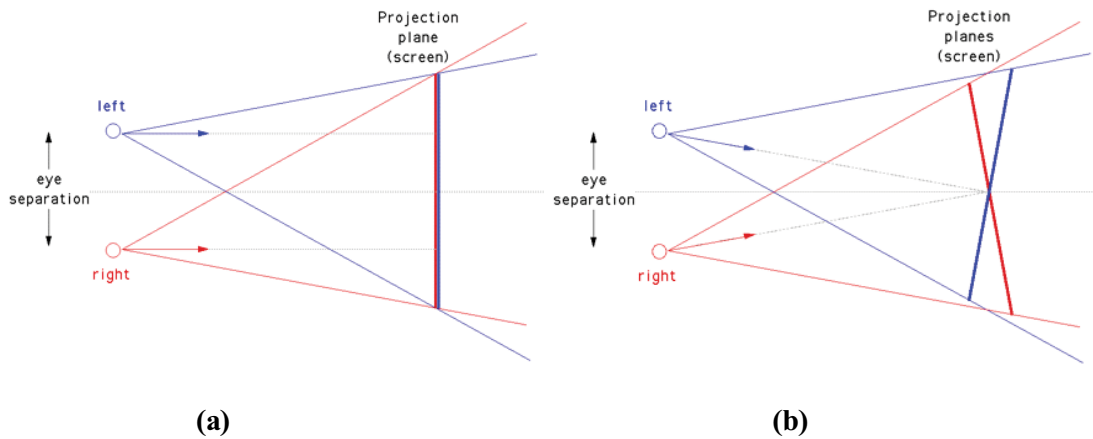






(b)

**Figure 2.5:** Stereo video capture, (a) stereo video camera, (b) Captured scenes



**Figure 2.6:** Dual camera configurations; (a) Parallel cameras, (b) Toed-in cameras

The 2D-to-3D conversion algorithms can be employed to transform 2D video into 3D video sequences. For instance, existing movies can be viewed as a novel stereoscopic film [13] [14]. In principle, 2D-to-3D conversion algorithms derive a depth map sequence from a 2D still image sequence. According to [15], the depth estimation techniques such as depth from motion and structure from motion will convert only a limited amount of the monoscopic video into 3D video automatically. Therefore, novel 2D-to-3D conversion methods are necessary with a limited manual intervention in order to support off-line and real-time conversion of 2D video into 3D video. The semi-automatic methods/algorithms developed by Dynamic Digital Depth Research Pty Ltd and Philips to recover the depth map of a monoscopic video are presented in [15] and [16] respectively.

The latest addition to the 3D capturing technology is the depth/range cameras. They simultaneously capture a colour image sequence and associated per-pixel depth image sequences of a scene. The 3D camera utilizes a light pulse to measure the relative depth of the objects in the scene (see Figure 2.7 (a)). Figure 2.7 (b) shows the internal architecture of the High Definition (HD) three-dimensional camera developed by NHK Laboratories Japan [17]. The Zcam™ [18] and Axi-vision [19] 3D cameras are two commercially available 3D depth/range cameras, which are developed by 3DV systems and NHK respectively. Moreover, these products are also available as add-ons for existing video capturing devices.

The snapshot of a scene captured with a 3D camera is given in Figure 2.8. The depth map sequence has similar spatio-temporal resolution as the colour image sequence. The depth images can be stored in 8 bit gray values, where gray value 0 specifies the furthest value (i.e. away from camera) and the gray level 255 specifies the closest value (i.e. closer to the camera). In order to translate this depth data representation to real, metric depth values and to support different image sequences with different depth characteristics, the gray levels are normalized into two main depth clipping plains namely;

- The near clipping plane  $Z_{near}$  (gray level 255), the smallest metric depth value  $Z$
- The far clipping plane  $Z_{far}$  (gray level 0), the largest metric depth value  $Z$ .



"I studied English for 16 years but...  
...I finally learned to speak it in just six lessons"  
Jane, Chinese architect

ENGLISH OUT THERE

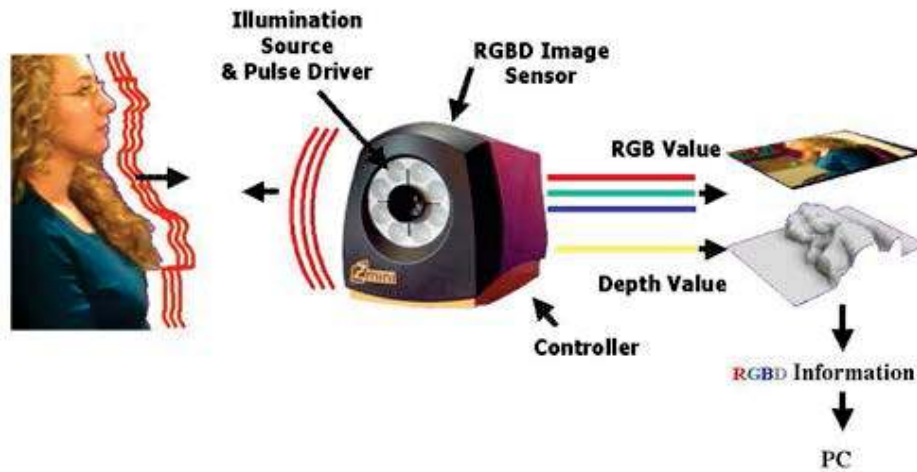
Click to hear me talking before and after my unique course download



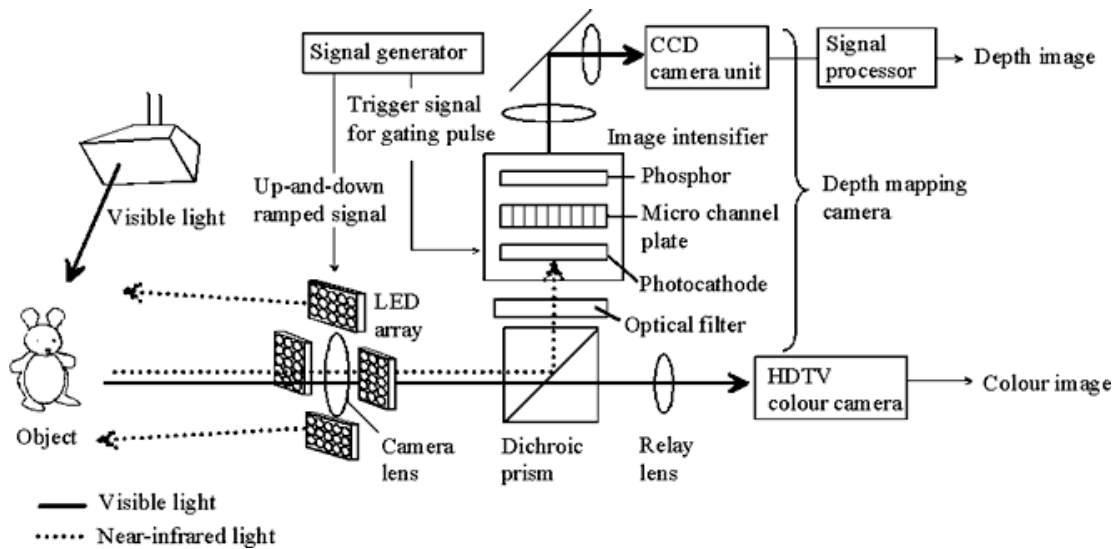
In case of linear quantization of depth, the intermediate depth values can be calculated using Equation 2.1.

$$Z = Z_{far} + v \left( \frac{Z_{near} - Z_{far}}{255} \right) \text{ with } v \in [0, \dots, 255] \quad \text{Equation 2.1}$$

where  $v$  specifies the respective gray level value.

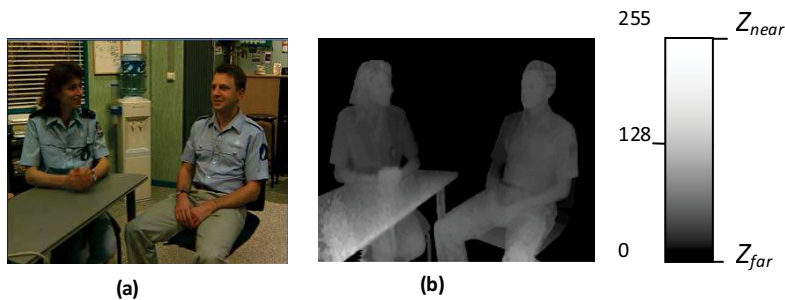


(a)



(b)

**Figure 2.7:** Depth/range camera; (a) Illustration of depth capture, (b) Internal architecture of a 3D camera [17]



**Figure 2.8:** Interview sequence; (a) Colour image, (b) Per-pixel depth image. The depth images are normalized to a near clipping plane  $Z_{near}$  and a far clipping plane  $Z_{far}$ .

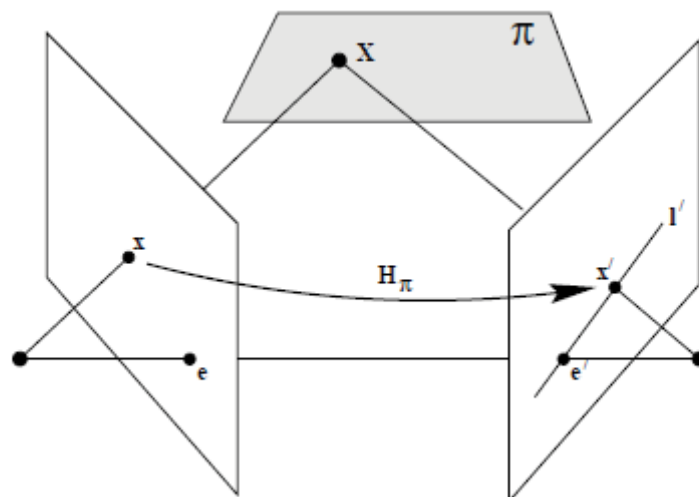
Depth-Image-Based Rendering (DIBR) can be utilized to render/synthesize two virtual views for the left and right eyes using the colour image sequence and the corresponding per-pixel depth information [20] [21]. This process can be employed in two major steps;

- The reprojection of original image point into a 3D space with the help of depth information
- The 3D space points are then projected into the image plane of the virtual camera.

In Computer Graphics (CG) this concept is known as 3D image warping. This concept is mathematically derived in subsection 2.2.1.1.

### 2.3 3D Image Warping

Figure 2.9 shows a system of two cameras and an arbitrary 3D space point  $X$  with the projection  $x$  and  $x'$  in the first and second virtual views respectively. All the image points are on the image plane of  $\pi$ .



**Figure 2.9:** A point  $x$  in one image is transferred via the plane  $\pi$  to a matching point  $x'$  in the second image. The epipolar line through  $x'$  is obtained by joining  $x'$  to the epipole  $e'$ . In symbols one may write  $x' = H_{\pi}x$  and  $l' = [e'] \times x' = [e'] \times H_{\pi}x = Fx$  where  $F = [e'] \times H_{\pi}$  is the fundamental matrix [22].



Under the assumption that the world coordinate system equals the camera coordinate system of the first camera, the two perspective projection equations will be;

$$\tilde{x} \cong AP_n \tilde{X} \quad \text{Equation 2.2}$$

$$\tilde{x}' \cong A'P_n H_f \tilde{X} \quad \text{Equation 2.3}$$

Where  $x$  and  $x'$  symbolize two 2D image points with respect to the 3D space point  $X$  in homogeneous notation. The symbol  $\cong$  denotes the 'equality up to a non-zero scale-factor' [22] [23]. The  $4 \times 4$  matrix  $H_f$  contains the transform matrix which converts the 3D space point from world coordinate system into the camera coordinate system of the second view. The  $H_f$  consists of two transform components namely rotation  $R$  and translation  $T$ . The  $3 \times 3$  matrices  $A$  and  $A'$  defines the intrinsic parameters of the first and second cameras respectively. The normalized perspective projection matrix is denoted by the  $3 \times 4$  identity matrix  $P_n$ .

The 3D space point  $X$  is still dependent on its depth value  $Z$ . Hence, Equation 2.2 can be rearranged into;

$$X = ZA^{-1}\tilde{m} \quad \text{Equation 2.4}$$

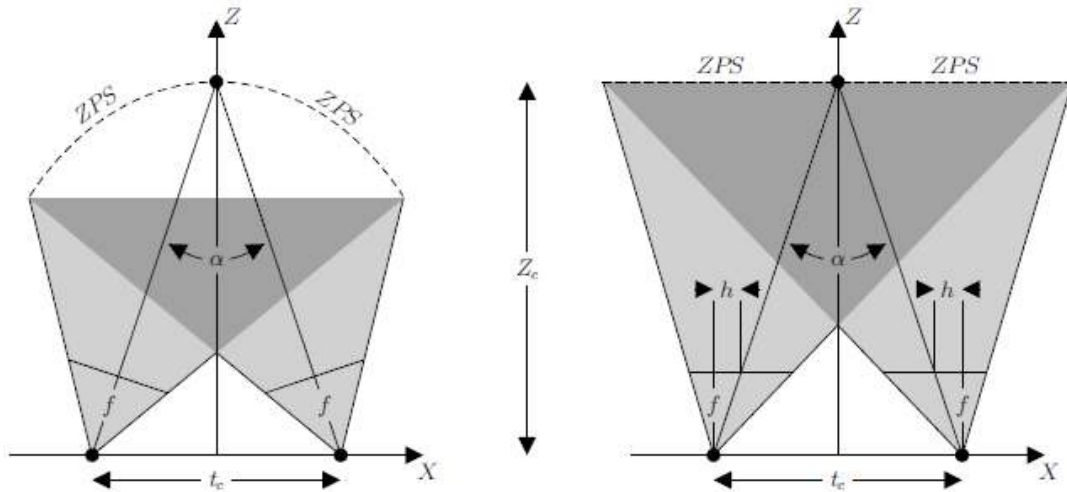
The depth dependent relationship between corresponding points in two perspective views of the 3D scene can be derived using Equations 2.3 and 2.4 and the outcome represents the classical disparity equation (see Equation 2.4).

$$Z'\tilde{m}' = ZA'RA^{-1}\tilde{m} + A't \quad \text{Equation 2.5}$$

This 3D image warping relationship can be utilized to render arbitrary novel views with respect to a known reference image. This requires the virtual camera position and orientation relative to the reference camera to be known with the intrinsic parameters of the virtual camera. Then if the depth values of the corresponding 3D space points are known for every pixel in the original image, novel views can be generated using Equation 2.5.

The virtual stereoscopic images can be generated through simplifying the 3D image warping technique (Equation 2.5) to represent horizontal parallax of two virtual camera positions. The relationship will be derived based on the two stereo camera configurations as shown in Figure 2.10. The both configurations can be distinguished based on how they achieve Zero-Parallax Setting (ZPS).

- Toed-in: The point of convergence at  $Z_c$  achieves through inward-rotation of the left and right cameras.
- Shift-sensor/parallel camera setup: A plane of convergence at  $Z_c$  is established through a shift  $h$  of the camera's CCD (Charged-Couple Device) sensor.



**Figure 2.10:** Different stereoscopic camera setups a). In the “toed-in” camera setup, a point of convergence at  $Z_c$  is established by a joint inward-rotation of the two cameras, b). In the shift-sensor camera setup, a plane of convergence at  $Z_c$  is established by a shift  $h$  of the camera’s CCD sensors. The  $t_c$  refers to the inter-axial distance between two cameras [24].

Excellent Economics and Business programmes at:



university of  
 groningen




“The perfect start  
 of a successful,  
 international career.”

**CLICK HERE**  
 to discover why both socially  
 and academically the University  
 of Groningen is one of the best  
 places for a student to be

[www.rug.nl/feb/education](http://www.rug.nl/feb/education)

The parallel camera setup is more suitable to be used with the DIBR technique, because all the signal processing steps going to be one-dimensional. With respect to the original view, the virtual cameras (i.e. left and right camera) are symmetrically displaced and their CCD sensors are shifted relative to the position of virtual camera lenses. This sensor shift can be mathematically formulated as a displacement of a camera's principal point  $c$  [23]. Therefore, intrinsic parameters of the virtual cameras are considered to be having similar intrinsic parameters of the original camera except the horizontal shift  $h$  of the respective principal point. This also can be formulated into an equation as follows;

$$A^* = \begin{bmatrix} \alpha_u & 0 & u_0 + h \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} = A + \begin{bmatrix} 0 & 0 & h \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Equation 2.6}$$

$A^*$  denotes the intrinsic parameters of either left (i.e.  $A^*$ ) or right (i.e.  $A^*$ ) virtual cameras.

With the assumption that the movement of the two virtual cameras are translational with respect to the reference camera (i.e. Rotation  $R = I$ , where  $I$  is the  $3 \times 3$  identity matrix) the 3D warping Equation 2.5 can be further simplified using the relationship derived in Equation 2.6;

$$A^*RA^{-1} = A^*A^{-1} = I + \begin{bmatrix} 0 & 0 & h \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{Equation 2.7}$$

Substituting Equation 2.5 with the simplified expression in Equation 2.7 the 3D warping equation can be written as;

$$Z^* \tilde{m}^* = Z \left( \tilde{m} + \begin{bmatrix} h \\ 0 \\ 0 \end{bmatrix} \right) + A^* t \quad \text{Equation 2.8}$$

With  $t_z = 0$ , the depth value of the 3D space point is same in the original camera (camera coordinate system of the original view) and the virtual camera (coordinate system of the virtual camera). Therefore,  $Z^* = Z$  and Equation 2.8 can be further reduced to;

$$\tilde{m}^* = \tilde{m} + \frac{A^* t}{Z} + \begin{bmatrix} h \\ 0 \\ 0 \end{bmatrix} \quad \text{with } t = \begin{bmatrix} t_x \\ 0 \\ 0 \end{bmatrix} \quad \text{Equation 2.9}$$

Then, the affine pixel positions  $(u,v)$  of each warped image can be calculated as;

$$\begin{aligned} u^* &= u + \Delta u \text{ and } v^* = v \\ &= u + \frac{\alpha_u t_x}{Z} + h \end{aligned} \quad \text{Equation 2.10}$$

The horizontal translational distance  $t_x$  is equal to the half of the inter-axial distance  $t_c$  (i.e. the average eye separation of humans, approximately 64 mm). The translational distance with the direction of the movement is;

$$t_x = \begin{cases} -\frac{t_c}{2} & : \text{Left-eye view} \\ +\frac{t_c}{2} & : \text{Right-eye view} \end{cases} \quad \text{Equation 2.11}$$

The amount of sensor shift  $h$  is dependent on the chosen convergence distance  $Z_c$ . When  $Z = Z_c$  the horizontal component  $u'$  of the simplified 3D warping Equation 2.10 is identical for both left and right views, i.e.  $u' = u''$ . Therefore, Equation 2.10 can be rewritten as;

$$h = -t_x \frac{\alpha_u}{Z_c} \quad \text{Equation 2.12}$$

where  $t_x$  is also defined by Equation 2.11.

Equations 2.10 and 2.11 can be utilized to render the virtual camera views of the parallel stereoscopic camera setup. The characteristics of the rendered virtual left and right views going to be affected by the choice of inter-axial distance  $t_c$ , the focal length  $f$  of the reference camera and the convergence distance  $Z_c$ . Table 2.1 shows how the 3D perception is affected due to these parameter settings. These effects can be attributed to the effects of real stereoscopic camera setup with the adjustments to their camera positions (e.g. inter-axial distance) and intrinsic parameters (e.g. focal length).

Parameter	+/-	Screen parallax	Perceived depth	Object size
Interaxial distance $t_c$	+	Increase	Increase	Constant
	-	Decrease	Decrease	Constant
Focal length $f$	+	Increase	Increase	Increase
	-	Decrease	Decrease	Decrease
Convergence distance $Z_c$	+	Decrease	Shift (forward)	Constant
	-	Increase	Shift (backward)	Constant

**Table 2.1:** Effects of different stereo camera setup parameters. Qualitative changes in screen parallax values, perceived depth and object size when varying the inter-axial distance  $t_c$ , the focal length  $f$  or the convergence distance  $Z_c$  of a “real” or “virtual” shift-sensor stereo camera setup [25].



### 2.3.1 The Advantages and Disadvantages of Colour Plus Depth Map Representation

The advantages of using colour plus depth map representation of stereoscopic video compared to the video generated with a stereo camera pair can be listed as follows:

- The 3D reproduction can be adjusted to different stereoscopic displays (e.g. auto-stereoscopic displays) and projection systems as the rendering happens at the receiver side.
- The 2D-to-3D conversion algorithms will generate more colour plus depth stereoscopic video and increase the timely availability of exiting stereoscopic materials.
- Head-Motion Parallax (HMP) could be supported which provides an additional stereoscopic depth cue. This format also limits the viewing angle of the stereoscopic video camera setup.
- Due to the smoothness characteristics of the real world objects the per-pixel depth information doesn't have high frequency components. Thus, the depth sequence can be efficiently compressed with existing compression standards [24] and will require only a limited space and bandwidth compared to the requirements of colour image sequence.
- The diminution of stereoscopic video sensation due to photometrical asymmetries (e.g. in terms of brightness, contrast or colour, between the left and the right eye) will be eliminated as this representation renders the virtual stereo views using the same colour image sequence.
- The depth reproduction can be adjusted at the receiver side based on user preferences (e.g. age, eye strain).
- This representation can be effectively used in 3D post production (e.g. augmenting external objects to the scene using object segmentation with the help of depth information).

However, the existing drawbacks of this representation, has led to several research findings which can be utilized to mitigate the effects of the monoscopic video plus depth map representations. The disadvantages of using this representation and the solutions come across are listed as follows.

- The quality of the rendered stereoscopic views depends on the accuracy of the per-pixel depth values of the original imagery. Therefore, the effects of compression and transmission of depth maps (e.g. introduced artefacts) on the perceived quality need to be thoroughly investigated.
- The visible objects for the rendered virtual left and right views may occlude from the original view. This phenomenon is also known as exposure and disocclusion in Computer Graphics (CG) [21]. This effect can be minimized using Layered Depth Images (LDI) where more than one pair of colour plus depth sequences is transmitted depending on the requirements of the expected quality [26]. However, this approach demands more storage and bandwidth to be used in communication applications. In addition, different hole-filling algorithms (e.g. linear interpolation of foreground and background colour, background colour extrapolation, mirroring of background colour information) can be utilized to recover the occluded areas of the original image sequence [24]. Moreover, the pre-processing/smoothing of depth maps (e.g. use of a Gaussian filter) will avoid this occlusion problem. However, this approach will lead to some geographical distortions of the rendered 3D video scene.
- Certain atmospheric effects (e.g. fog, smoke) and semi-transparent objects are difficult to handle with this approach at the moment.
- The processing overload (e.g. memory, processing power, storage requirements) at the receiver side is high compared to the reconstructing 2D video stream.

The monoscopic video plus depth map representation is widely utilized in research and standardization activities due to its simplicity and adaptability [27] [28] [29]. The ATTEST (Advanced Three-Dimensional Television System Technologies) project consortium is working on 3D-TV broadcast technologies using colour-depth sequences as the main source of 3D video [27]. Recently, ISO/IEC 23002-3 (MPEG-C part 3) finalized the standardization of video plus depth image representations/solutions in order to provide: interoperability of the content, flexibility regarding transport and compression techniques, display independence and ease of integration [28]. Moreover, JVT has identified multi-view video plus depth representation would be a potential candidate for free-view point applications [29]. Due to this wide usage in research and standardization activities, the research carried out in this book utilize the colour plus depth map 3D video representation. This selection would be also supported by the range of advantages associated with this scene representation. For example, the transmission of colour plus depth map would require fewer system resources (bitrate, storage) than the resource requirements for sending left and right views.

Four colour and depth map based stereoscopic video sequences namely Orbi, Interview, Break dance and Ballet are used in the experiments presented in this book. Figure 2.11 shows frames from the original scenes of these test sequences. The Orbi and Interview test video sequences are obtained using a depth/range camera (i.e.  $Z_{\text{cam}}$ ™ camera) are used in the experiments [30]. Orbi is a very complex sequence with camera motion and multiple objects, whereas Interview is a sequence captured with a static camera and featuring a stationary background. The resolution of these two sequences is 720×576 pixels which is the resolution of Standard Definition (SD) TV and original frame rate is 25 frames/s. The rest of the sequences (i.e. Break dance and Ballet) are obtained from the multi-view image sequences generated by the Interactive Visual Media group at Microsoft Research [31]. The fourth camera view and the corresponding depth map computed from stereo are utilized in this experiment [31]. Break dance sequence contains a highly dynamic break dancer in the foreground and a number of supporters with limited motion in the background. In contrast to the Break dance test sequence, Ballet occupies a stationary observer in the foreground and a Ballet dancer operating behind the foreground observer. Both sequences are captured using a stationary camera. The resolution and original frame rate of these two sequences are 1024×768 and 15 frames/s respectively. Due to the use of different colour and depth map sequences (e.g. resolution, frame rate) the results of the experiments will be applicable across most of the application scenarios. Moreover, the findings will be common for all colour plus depth map video representations regardless of the way the material are captured. The issues associated with compression, transmission, display and quality evaluations of this stereoscopic representation are discussed in the following sections.

## LIGS University

based in Hawaii, USA

is currently enrolling in the  
Interactive Online **BBA, MBA, MSc,**  
**DBA and PhD** programs:

- ▶ enroll **by October 31st, 2014** and
- ▶ **save up to 11%** on the tuition!
- ▶ pay in 10 installments / 2 years
- ▶ Interactive Online education
- ▶ visit [www.ligsuniversity.com](http://www.ligsuniversity.com) to find out more!

Note: LIGS University is not accredited by any nationally recognized accrediting agency listed by the US Secretary of Education. More info [here](#).







**Figure 2.11:** Original colour and depth based 3D test sequences; (a) Orbi, (b) Interview, (c) Break-dance, and (d) Ballet