Identification of stereo rig alignment error based on vertical disparity map

Sandeep Katragadda

KiwiSecurity Software GmbH,
Guglgasse 15, 1110, Vienna, Austria

s.katragadda@kiwisecurity.com

ABSTRACT

Poor quality of 3D video content can lead to headache, blurry vision and overall exhaustive experience for the viewer. To ensure quality and comfortable 3D experience for the end consumer, common production errors must be detected and corrected. Vertical disparity is one of these distortions and is caused by improper stereo-camera setup. This paper aims at identifying the possible rotational and placement errors that cause the vertical disparity. An estimation of these errors is necessary to produce good quality 3D content. According to my knowledge, there exists no method to identify rig alignment errors without the knowledge of camera setup parameters and this work is the first step in that direction. Feature point detection has proven to be an interesting approach to find vertical disparity present in the given stereo image pair. In this work feature extraction techniques such as SIFT, SURF and Harris features are efficiently used to compute reliable and robust vertical disparity patterns. This paper classifies vertical disparity patterns according to rig errors. If the vertical disparity values are too small or ambiguous to be identified by pattern analysis, this paper uses graphical analysis that highlights the relationship between the total vertical disparity and the contribution of each possible error to the total. Experimental results show that the proposed approach identifies the reason behind the presence of vertical disparity of a stereo image pair.

Keywords
Stereoscopic 3D, feature matching, vertical disparity, epipolar constraint, RANSAC.

1 INTRODUCTION

Experiencing 3D is one of the most important perception features of human kind. Several depth cues are used by human visual system to perceive depth and to experience 3D. Stereoscopic video systems simulate the human binocular vision [Sammons92]. Binocular depth cues are observed by both eyes. They include retinal disparity in the two views of a scene (one from each eye) resulted from inter-ocular distance, and convergence due to the inward movement of both eyes to focus at a point of interest. If the cues are not reproduced properly in the stereo content, it results in a bad user experience [Perek16, Tam11, Hodges91]. Poor quality of stereoscopic video content results in visual fatigue [IJsselsteijn00, ATSC11]. Consequences of visual fatigue on human 3D perception might be eye strain, watery eyes, nausea, head aches, focusing difficulty and blurred vision. Therefore, stereographers and

camera teams have to take into account a variety of conditions, guidelines and rules right from the beginning of the production chain to create good stereoscopic video content [Zilly11, Knorr12]. It includes accurate rigging and calibration of the stereo cameras, good adjustment and matching of electronic and optical camera parameters as well as the adaption of the stereo baseline to the depth structure of the scene content.

When there is an error in the alignment of the stereo rig setup, vertical disparity can be observed in the image pair. Vertical disparity is defined as the vertical displacement between corresponding pixels in the left and right images. Vertical disparity in the stereo content is result of distortions. There are other distortions such as keystone distortion, depth plane curvature, barrel distortion and pin cushion distortion caused by other factors. One of the results of vertical disparity is vertical parallax which can cause eye strain, visual fatigue, confusion and loss of stereopsis. The alignment errors in the stereo rig setup include rotational errors, translational errors or both. The rotational errors include relative roll, relative pan and relative tilt between the cameras. The translational errors include relative vertical shift and relative forward or backward shift between the cameras. One more error is zoom difference between the two cameras. It is caused when the lenses used in the two cameras have different focal length. Converged

(toed-in) camera setup and parallel camera setup are the two stereoscopic camera configurations used in 3D capturing [Woods93]. In the toed-in setup, the two cameras are angled towards each other resulting in a finite convergence distance (the distance at which the two camera images coincide in the centre of the stereoscopic display). The setup causes keystone distortion and vertical disparities which can be corrected using stereo image rectification [Faugeras93] during post processing. In the parallel setup, the optical axes of the two cameras are parallel so the convergence distance is infinity. Horizontal Image Translation (HIT) is done during post processing to place the point of convergence wherever desired. In the ideal case, the parallel setup does not produce vertical disparity.

This paper mainly focuses on vertical disparity caused by stereo rig alignment errors in the parallel setup and proposes a method to identify rig alignment errors based on vertical disparity in the given stereo image pair. These errors include relative rotation, relative translation and relative zoom difference. It is assumed that no lens distortion is present in the given image pair and the image pair is not rectified after capturing. The steps include reliable estimation of the vertical disparity and finding what kind of stereo rig error is causing this vertical disparity. The main contributions are (1) classification of possible vertical disparity patterns according to the stereo rig errors, and (2) identification of rig alignment error using vertical disparity map.

The paper is organised as follows. Section 2 briefly presents the state of the art on stereo rig alignment error identification. Section 3 presents the required mathematical background of stereo capturing. Section 4 proposes the vertical disparity based alignment error identification method. Section 5 presents the experimental results. Finally, Section 6 concludes the paper.

2 STATE OF THE ART

For a calibrated system, there is a lot of work presented in the literature to identify rig alignment errors. Behrooz Kamgar-Parsi et al. [Kamgar-Parsi88] computed small rotational errors caused by mechanical difficulties in the stereo rig setup. They proved that an error of 0.5 degrees in pan angle can result in an error of 4 meters in depth estimation which emphasizes the importance of accurate alignment. They assumed that both cameras are calibrated, i.e., the focal length, spatial digitisation of image plane and lens distortion are known. Behrooz's work also emphasizes that the points selected from a specific region cannot cover all properties of relative rotations. Hence, it is important to select point correspondences from different regions of the image. In case of uncalibrated camera systems, identification of stereo rig error is still a challenge. Richard I. Hartley [Hartley92] and Frederik Zilly
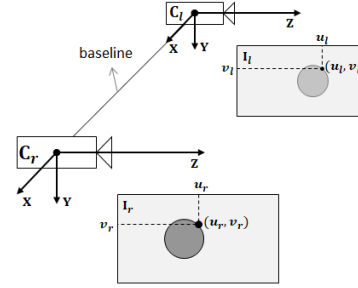

Figure 1: Stereo camera geometry in parallel setup. Key - C_l and C_r are optical centres of left and right cameras respectively. I_l and I_r are the left and right image planes respectively. (u_l, v_l) and (u_r, v_r) are the pixel locations of a real point on left and right image planes respectively.

et al. [Zilly10] proposed methods to compute the alignment errors from an uncalibrated stereo by using some assumptions on intrinsic parameters. Hartley did a lot of research in multiple view geometry [Hartley04]. He proposed two algorithms to calculate the relative placement of the cameras in the stereo rig [Hartley92]. One algorithm deals with the calibrated camera system and the other deals with the uncalibrated camera system. In case of calibrated system, it is assumed that everything is known except relative positions and relative orientations. In case of uncalibrated system, it is assumed that the two cameras are of the same type (having the same intrinsic parameters and the same lenses) and the principle point is at the image centre, the focal lengths of the two cameras, their relative placement and orientation are determined. This approach results in imaginary values when the assumptions fail to satisfy. Hence the result is not guaranteed in all cases. Frederik Zilly et al. proposed a new method [Zilly10] that computes rectifying homographies using sparse stereo correspondences from an uncalibrated stereo. This paper uses this approach to classify the disparity patterns caused by different alignment errors and to identify the stereo rig errors.

3 MATHEMATICAL BACKGROUND

Consider a parallel stereo setup shown in Figure 1 and 2. The line joining the optical centres C_l and C_r of the left and right cameras respectively is called baseline. Consider a point P in space whose perspective projection is at point $p_l = (u_l, v_l)$ on the image plane I_l of left camera and at point $p_r = (u_r, v_r)$ on the image plane I_r of right camera. The plane connecting P, C_l and C_r is called epipolar plane. The plane intersects the image planes I_l and I_r in lines called epipolar lines. Let $P_r = [X_r\ Y_r\ Z_r]$ represent the same point P in the right camera coordinate system. The perspective projection of the point P_r on the image plane I_r is at $(f_r X_r / Z_r, f_r Y_r / Z_r)$ where f_r is the focal length of the right camera. To obtain this image location in terms of pixel values, intrinsic parameters of the camera have to
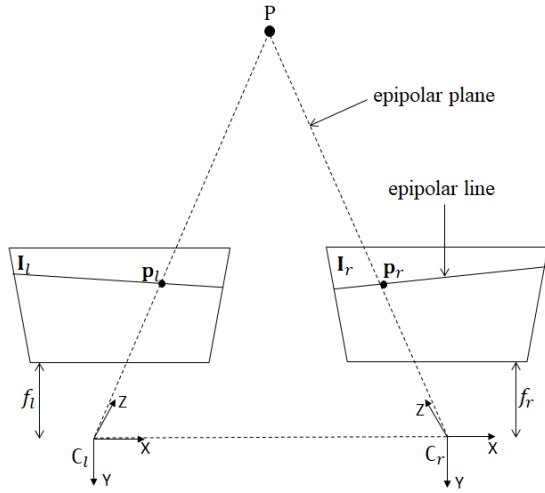
Figure 2: Epipolar geometry. Key - $\mathbf{C}_l$ and $\mathbf{C}_r$ are optical centres of left and right cameras respectively. $\mathbf{I}_l$ and $\mathbf{I}_r$ are the left and right image planes respectively. $\mathbf{P}$ is a point in the scene. $\mathbf{p}_l$ and $\mathbf{p}_r$ are the pixel locations of $\mathbf{P}$ on left and right image planes respectively. f_l and f_r are the focal lengths of the left and right cameras respectively.

be considered. They include principle point [Clarke98], pixel dimensions [Lenz88] and skew coefficient describing non-rectangular pixels. If (c_r^x, c_r^y) is the pixel coordinate of the principle point, s_r^x and s_r^y represent effective pixel sizes along the horizontal and vertical directions of the image respectively [Lenz88], the 3D point $\mathbf{P}_r$ to 2D point $\mathbf{p}_r$ transformation is [Hartley04]

$$\begin{aligned} u_r &= (f_r X_r/Z_r)/s_r^x + c_r^x, \\ v_r &= (f_r Y_r/Z_r)/s_r^y + c_r^y. \end{aligned} \tag{1}$$

An upper triangular matrix describing these intrinsic properties of the right camera holding the parameters such as principle point (c_r^x, c_r^y), pixel dimensions f_r/s_r^x and f_r/s_r^y, and skew s_r is defined as

$$\mathbf{K}_r = \begin{bmatrix} f_r/s_r^x & s_r & c_r^x \\ 0 & f_r/s_r^y & c_r^y \\ 0 & 0 & 1 \end{bmatrix}. \tag{2}$$

Assuming zero skew, equation 1 can be written as

$$\begin{bmatrix} \mathbf{p}_r \\ 1 \end{bmatrix} = \mathbf{K}_r \hat{\mathbf{P}}_r, \tag{3}$$

where $\hat{\mathbf{P}}_r = [X_r/Z_r \; Y_r/Z_r \; 1]^\mathsf{T}$ represents the projection on the right image plane (the superscript T refers to the transpose operation). Similarly, $\mathbf{p}_l$ is image point (u_l, v_l) of the same world point $\mathbf{P}$ whose 3D location is $\mathbf{P}_l$ in the left camera coordinate system and $\hat{\mathbf{P}}_l$ represents the projection on the left image plane.

Epipolar geometry describes the geometric relationship between a pair of stereo images [Boufama95]. Let the left camera coordinate system be the reference coordinate system, $\mathbf{R}$ be the 3×3 matrix (the rotation matrix) defining the relative rotation of the right camera

with respect to the left camera and $\mathbf{T}$ be the 3×1 vector (the translation vector) defining the relative translation of the right camera with respect to the left camera. Then the relation between $\hat{\mathbf{P}}_r$ and $\hat{\mathbf{P}}_l$ is $\hat{\mathbf{P}}_r = \mathbf{R}(\hat{\mathbf{P}}_l - \mathbf{T})$[Andrea00]. From the plane equation of two vectors, the equation of epipolar plane is

$$\begin{aligned} (\hat{\mathbf{P}}_l - \mathbf{T})^\mathsf{T} . (\mathbf{T} \times \hat{\mathbf{P}}_l) &= 0, \\ (\mathbf{R}^\mathsf{T} \hat{\mathbf{P}}_r)^\mathsf{T} [\mathbf{T}]_\times \hat{\mathbf{P}}_l &= 0. \end{aligned} \tag{4}$$

Here $[\mathbf{T}]_\times$ is the skew-symmetric matrix of vector $\mathbf{T}$. From equation 3 and equation 4,

$$\begin{bmatrix} \mathbf{p}_r & 1 \end{bmatrix} \mathbf{F} \begin{bmatrix} \mathbf{p}_l \\ 1 \end{bmatrix} = 0, \tag{5}$$

where $\mathbf{F} = (\mathbf{K}_r^{-1})^\mathsf{T} \mathbf{R}[\mathbf{T}]_\times \mathbf{K}_l^{-1}$ is called fundamental matrix between the pair of stereo images. Equation 5 is called Epipolar constraint. This maps points from one image to corresponding epipolar lines in the other image. The fundamental matrix $\mathbf{F}$ encodes the information of both extrinsic ($\mathbf{R}$ and $\mathbf{T}$) and intrinsic parameters ($\mathbf{K}_l$ and $\mathbf{K}_r$). A stereo camera system is said to be calibrated if its intrinsic parameters and the extrinsic parameters are known. If some of the parameters are known it is called semi calibrated or partially calibrated. If nothing about these parameters is known the system is an uncalibrated system.

In the ideal parallel stereo setup (parallel Z-axis of the cameras) case, the X-axis of each camera is colinear with the base line and hence projects the point $\mathbf{P}$ on to the two image points $\mathbf{p}_r$ and $\mathbf{p}_l$ on the same scan line making the vertical disparity $v_r - v_l$ zero. The distance between the point $\mathbf{P}$ and the stereo camera is proportional to the horizontal disparity $u_r - u_l$.

4 PROPOSED METHOD

Let α_x, α_y and α_z are the relative rotation angles of the right camera with respect to the left camera around X, Y and Z axes respectively, and t_x, t_y and t_z are the relative translation of right camera with respect to the left camera along X, Y and Z axes respectively. This work assumes that (1) the relative rotation angles are small, i.e. $\alpha_x << 1$, $\alpha_y << 1$ and $\alpha_z << 1$, and the rotation matrix $\mathbf{R}$ can be approximated as

$$\hat{\mathbf{R}} = \begin{bmatrix} 1 & -\alpha_z & \alpha_y \\ \alpha_z & 1 & -\alpha_x \\ -\alpha_y & \alpha_x & 1 \end{bmatrix}, \tag{6}$$

(2) relative translation in Y and Z directions are small compared to that in X-direction (baseline length), i.e. $\hat{t}_y = t_y/t_x << 1$ and $\hat{t}_z = t_z/t_x << 1$, so without affecting Equation 5, $\mathbf{T}$ can be taken as $\hat{\mathbf{t}} = [1 \; \hat{t}_y \; \hat{t}_z]^\mathsf{T}$ and

$$[\hat{\mathbf{t}}]_\times = \begin{bmatrix} 0 & -\hat{t}_z & \hat{t}_y \\ \hat{t}_z & 0 & -1 \\ -\hat{t}_y & 1 & 0 \end{bmatrix}, \tag{7}$$
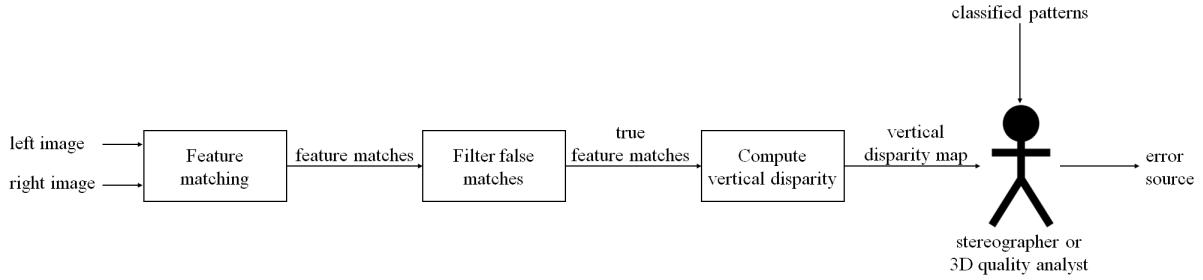
Figure 3: Block diagram showing the steps in the proposed classification based rig alignment error identification approach.

(3) pixels are square shaped meaning the skew of the cameras is zero (i.e. $s_r = s_l = 0$), $s_r^x = s_r^y$ and $s_l^x = s_l^y$, (4) focal lengths' difference α_f is very small such that $\hat{f}_r = (1 - \alpha_f)/\hat{f}_l$ where $\hat{f}_l = f_l/s_l^x$ and $\hat{f}_r = f_r/s_r^x$ are the focal lengths of left and right cameras (in pixels) respectively.

By substituting the assumptions in Epipolar constraint (equation 5) and ignoring the second and higher order terms, the total vertical disparity $v_r - v_l$ corresponding to the point **P** can be written as a function of $(\hat{f}_l, \alpha_f, \alpha_x, \alpha_y, \alpha_z, \hat{t}_y, \hat{t}_z)$ as [Zilly10]

$$v_r - v_l = \underbrace{\hat{t}_y(u_r - u_l)}_{Y-shift} + \underbrace{\frac{\hat{t}_z}{\hat{f}_l}(u_l v_r - u_r v_l)}_{Z-shift} + \underbrace{\alpha_f v_r}_{zoom\ difference}$$
$$\underbrace{- \alpha_x \hat{f}_l - \frac{\alpha_x}{\hat{f}_l} v_l v_r}_{tilt} + \underbrace{\frac{\alpha_y}{\hat{f}_l} u_r v_l}_{pan} + \underbrace{\alpha_z u_r}_{roll}.$$

$$(8)$$

It can be seen from the equation that different errors produce different vertical disparity patterns in the image. The following section shows the disparity patterns caused by some errors. If stereographers or quality analysts are familiar with these patterns, they can identify the error causing the pattern in a given stereo image pair. Vertical disparity map can be obtained by using sparse features matching algorithms such as SIFT [Lowe04], SURF [Bay08] and Harris features. Figure 3 shows the block diagram of the proposed pattern classification based error identification approach.

4.1 Classification of disparity patterns

It has been shown that without the knowledge of the exact focal length of each camera in the stereo rig, computation of exact relative angles and relative positions is not possible. Moreover the focal length changes with changes in zoom. However, irrespective of the focal length value used to capture a scene, properties of the vertical disparity map are unique for each kind of error. Using equation 8, different stereo rig errors are classified according to the properties of the vertical disparity map. These properties can be used to identify the error sources based on the vertical disparity pattern present in the given stereo image pair. This section discusses the
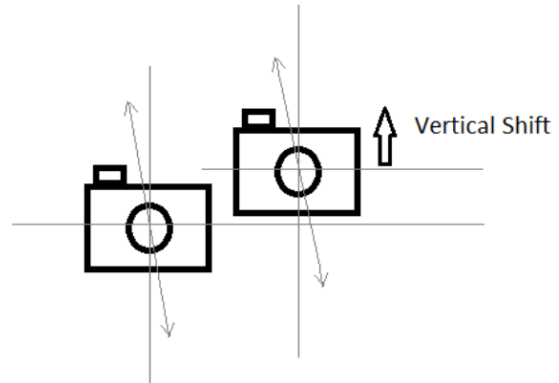

Figure 4: Cameras are at different heights

vertical disparity map properties for all possible stereo rig errors.

4.1.1 Relative translation in Y-direction

If the two cameras in a stereo setup are placed at different heights (X-axes of the cameras are separated vertically as shown in Figure 4), one image is shifted upward or downward relative to the other depending on the vertical shift direction. If the right camera is placed above the left camera, all left image features are moved downwards in the right image. See Figure 5(a) where the dashed border shows the right image and the thick border shows the left image. The vertical disparity is represented using arrows. Each arrow's head and tail locate at the row value (y- coordinates) of corresponding feature pixel in the right and the left images respectively. Similarly when the right camera is placed below the left camera, all left image features are moved upwards in the right image as shown in the Figure 5(b).

If vertical translation is the only error present in a stereo rig, equation 8 can be reduced to

$$v_r - v_l = \hat{t}_y(u_r - u_l), \qquad (9)$$

where $u_r - u_l$ is horizontal disparity representing the depth. The properties of this error are: (1) vertical disparity is directly proportional to the horizontal disparity (depth) magnitude and (2) the value of vertical disparity is the same for all points located at a particular depth. These vertical disparity patterns can be seen in Figure 6.
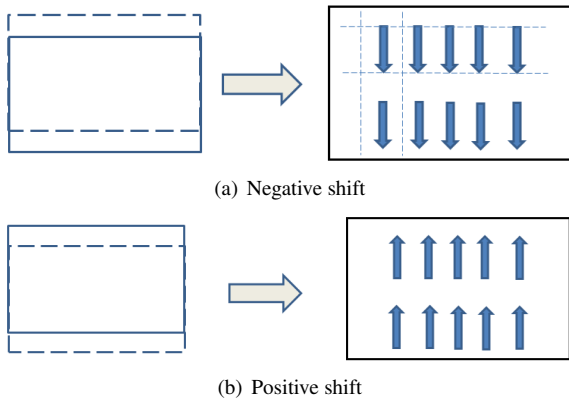
(a) Negative shift

(b) Positive shift

Figure 5: Placement error along Y-axis of left camera (thick rectangle) and right camera (dashed rectangle), and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.
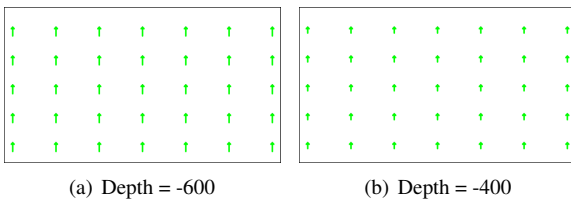
(a) Depth = -600   (b) Depth = -400

Figure 6: Vertical disparity patterns at different depths ($u_r - u_l$) when the cameras are at different heights ($\hat{t}_y = 0.1$).

4.1.2 Relative translation in Z-direction

If one camera is placed a little forward or backward (along Z-axis) to the other camera, it results in vertical disparity. Figure 7 helps to visualise this type of setup error. If focal lengths of the two cameras are the same, a disparity pattern as shown in Figure 8(a) and Figure 8(b) can be observed. If translation in Z-direction is the only error present in a stereo rig, equation 8 can be reduced to

$$v_r - v_l = \frac{\hat{t}_z}{\hat{f}_l}(u_l v_r - u_r v_l).$$ (10)

The properties of this error are: (1) as the vertical disparity depends on u_r and u_l values, the disparity map varies with the depth, (2) the value of vertical parallax increases or decreases from the image top edge to the bottom edge, (3) the vertical disparity decreases as the depth increases and (4) vertical parallax is symmetrical about the row at the image centre. These vertical disparity patterns can be seen in Figure 9. Disparity due
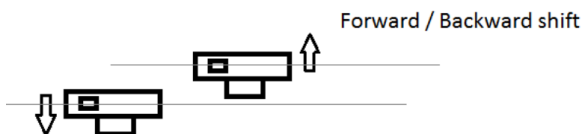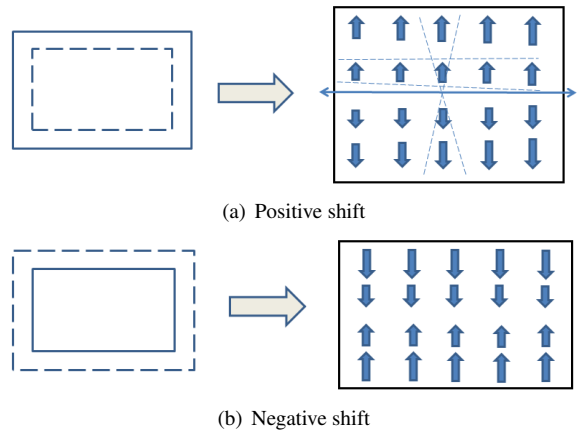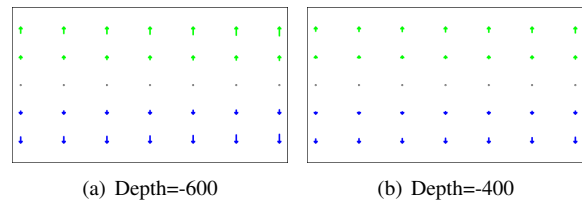

Figure 7: Cameras with a shift along the optical axis

(a) Positive shift

(b) Negative shift

Figure 8: Placement error along Z-axis of left camera (thick rectangle) and right camera (dashed rectangle), and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.

(a) Depth=-600   (b) Depth=-400

Figure 9: Vertical disparity patterns at different depths ($u_r - u_l$) when the cameras are separated in Z-direction ($\hat{t}_z = 0.1$).

to translation in Z-direction decreases with increasing depth and becomes zero at infinity (i.e. at depth=0).

4.1.3 Zoom difference

Another possible alignment error is focal length error. If two cameras use slightly different focal length values (different zoom), it shows similar vertical disparity pattern as previous (shift in Z-direction). See Figure 10(a) and Figure 10(b) to visualise the effect. If relative zoom is the only error present in a stereo rig, equation 8 can be reduced to

$$v_r - v_l = \alpha_f v_r.$$ (11)

The properties of relative zoom error are: (1) the disparity map is independent of depth information. This is the key point that differentiates this error from translational error in Z- direction where the vertical parallax varies with depth, (2) the amount of vertical parallax increases or decreases from the image top to the bottom and (3) The magnitude of vertical disparity is constant in a row. These vertical disparity patterns can be seen in Figure 11.

4.1.4 Relative rotation about X-axis (Tilt)

If one camera is rotated about X-axis relative to the other (as shown in Figure 12), features move up or down in the other camera's view. See Figure 13(b) and

(a) Positive zoom difference
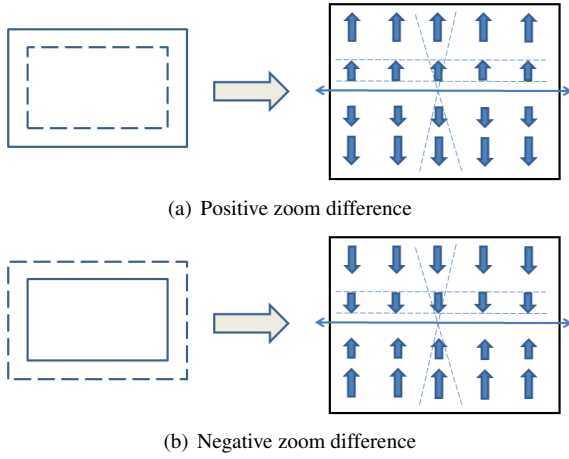

(b) Negative zoom difference

Figure 10: Zoom error representation with left camera (thick rectangle) and right camera (dashed rectangle), and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.
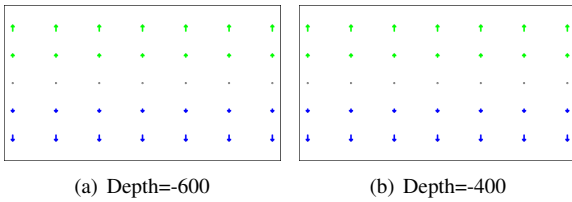

(a) Depth=-600   (b) Depth=-400

Figure 11: Vertical disparity patterns at different depths $(u_r - u_l)$ when there is different zoom in camera pair $\alpha_f = 0.1$.

Figure 13(a) to visualise the effect. If relative tilt angle is the only error present in a stereo rig, equation 8 can be reduced to

$$v_r - v_l = -\alpha_x \hat{f}_l - \frac{\alpha_x}{\hat{f}_l} v_l v_r. \qquad (12)$$

The properties of relative tilt error are: (1) the disparity map is independent of depth information, (2) the amount of vertical parallax increases or decreases from the image centre to the upper and to the lower edges and (3) if the focal length is very large, the component $-\frac{\alpha_x}{\hat{f}_l} v_l v_r$ can be ignored. In that case vertical disparity is uniform through out the image. These vertical disparity patterns can be seen in Figure 14.
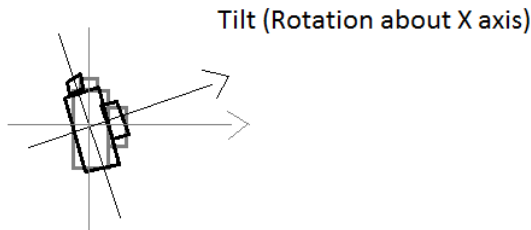

Figure 12: Cameras with relative tilt angle.
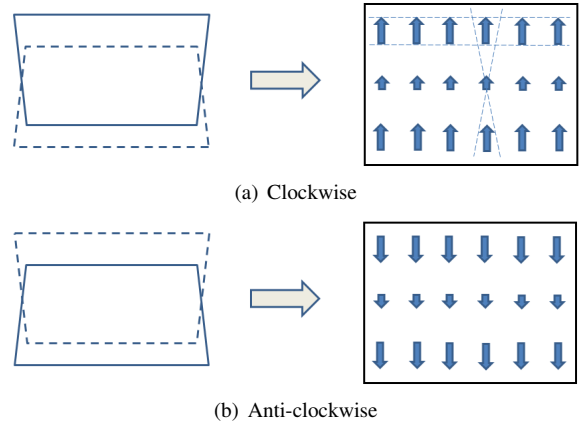
(a) Clockwise

(b) Anti-clockwise

Figure 13: Relative rotation along X-axis error of left camera (thick rectangle) and right camera (dashed rectangle) and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.
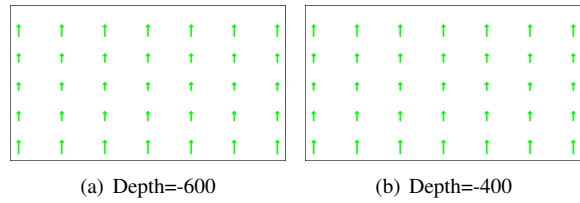

(a) Depth=-600   (b) Depth=-400

Figure 14: Vertical disparity patterns at different depths when the relative tilt angle between the cameras $\alpha_x = 0.1$.

4.1.5 Relative rotation about Y-axis (Pan)

If one camera is rotated about Y-axis relative to the other (as shown in Figure 15), it is called toed-in / toed-out setup. It results in key stone distortion. In this case, the vertical disparity can be observed clearly at the corners than at the image centre. Opposite corners exhibit similar disparity pattern while adjacent corners show opposite directions. See Figure 16(b) and Figure 16(a) to visualise the effect. If relative pan angle is the only error present in a stereo rig, equation 8 can be reduced to

$$v_r - v_l = \frac{\alpha_y}{\hat{f}_l} u_r v_l. \qquad (13)$$

The properties of the relative pan angle are: (1) the disparity map is independent of depth information, (2) the direction and magnitude of the vertical parallax is the same on the opposite corners, (3) neighboring corners
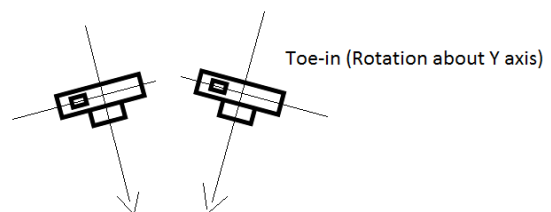

Figure 15: Cameras with relative pan angle.
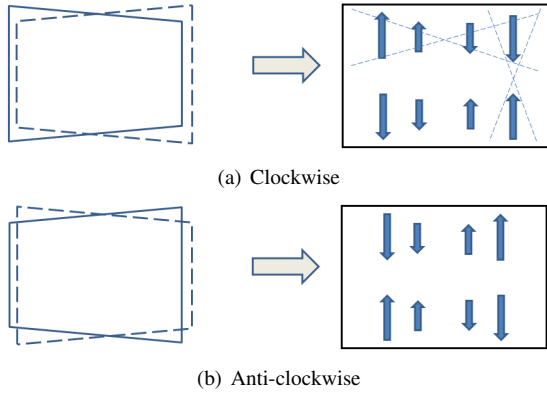
(a) Clockwise

(b) Anti-clockwise

Figure 16: Relative rotation along Y-axis error of left camera (thick rectangle) and right camera (dashed rectangle) and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.

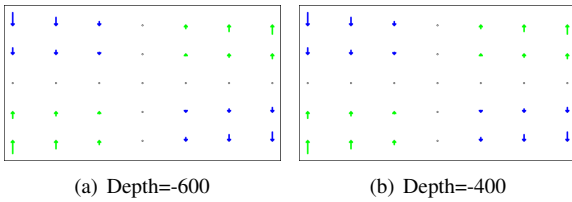(a) Depth=-600   (b) Depth=-400

Figure 17: Vertical disparity patterns at different depths when the relative pan angle between the cameras $\alpha_y = 0.1$.

differ in the direction of vertical disparity but the magnitude of the vertical disparity is the same and (4) the value of vertical disparity increases from the top edge to the bottom edge on one side and decreases on the other side of the image central column. These vertical disparity patterns can be seen in Figure 17.

4.1.6   Relative rotation about Z-axis (Roll)

If one camera is rotated about Z-axis relative to the other (as shown in Figure 18), the features on one side of the central column move up and on the other side move down. See Figure 19(a) and Figure 19(b) to visualise the effect. If relative roll angle is the only error present in a stereo rig, equation 8 can be reduced to

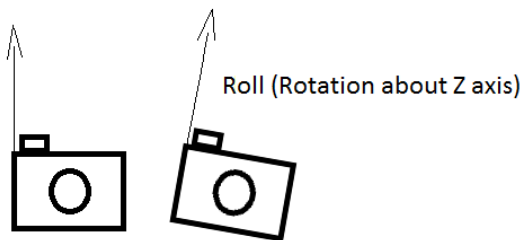$$v_r - v_l = \alpha_z u_r. \qquad (14)$$

Figure 18: Cameras with relative roll angle.

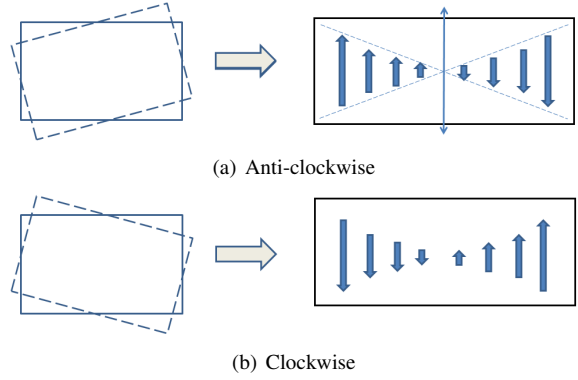(a) Anti-clockwise

(b) Clockwise

Figure 19: Relative rotation along Z-axis error of left camera (thick rectangle) and right camera (dashed rectangle) and the corresponding vertical disparity pattern. The distance between dashed lines in pattern represents the magnitude of disparity.
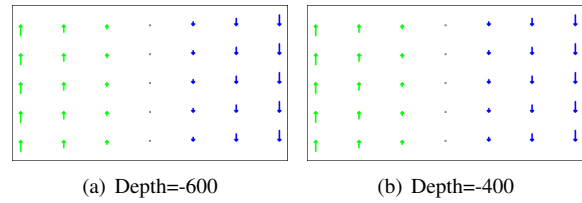

(a) Depth=-600   (b) Depth=-400

Figure 20: Vertical disparity patterns at different depths when the relative roll angle between the cameras $\alpha_z = 0.1$.

The properties of the relative roll angle are: (1) the disparity map is independent of depth information, (2) the value of vertical parallax increases or decreases from the left edge to the right edge of the image (in a row) and (3) the vertical disparity pattern is symmetrical about central column of the image. These vertical disparity patterns can be seen in Figure 20.

4.1.7   Multiple errors (Tilt and Roll)

The patterns can be classified also for combination of the errors. Here, only the tilt and roll combination is presented. If relative tilt and relative roll are present in a stereo rig, it results in disparity patterns as shown in Figure 21. Equation 8 can be reduced to

$$v_r - v_l = \alpha_z u_r - \hat{f}_l \alpha_x - \frac{\alpha_x}{\hat{f}_l} v_l v_r. \qquad (15)$$

The properties of the resulting vertical disparity map are: (1) the disparity map is independent of depth (prop-
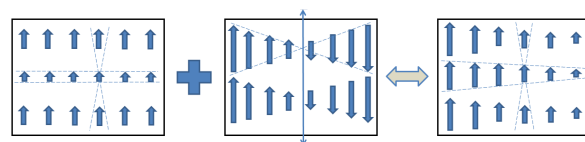

Figure 21: Resulting vertical disparity pattern when relative rotations along X-axis (tilt) and Z-axis (roll) are present simultaneously. The distance between dashed lines in pattern represents the magnitude of disparity.
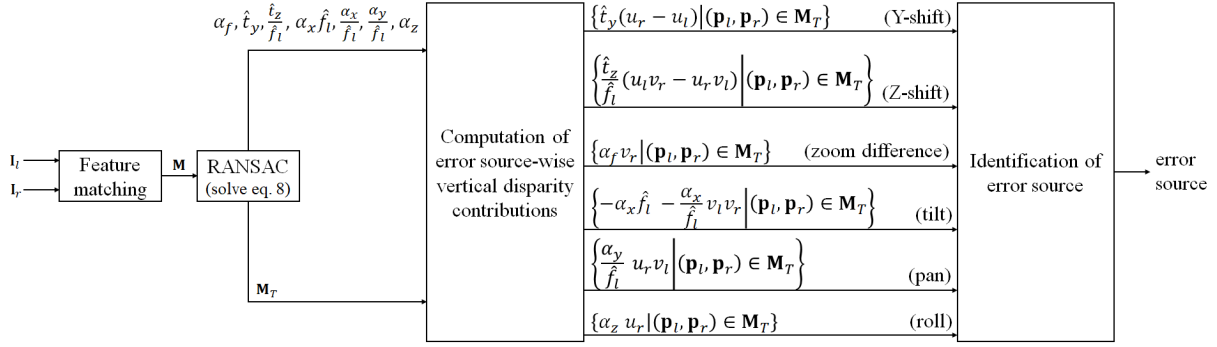
Figure 22: Block diagram showing the steps in the proposed mathematical approach. Key - $\mathbf{I}_l$: left image, $\mathbf{I}_r$: right image, $\mathbf{p}_l = (u_l, v_l)$: a point on the left image, $\mathbf{p}_r = (u_r, v_r)$: a point on the right image, $\mathbf{M}$: set of feature point matches, $\mathbf{M}_T \subseteq \mathbf{M}$: set of true feature point matches, α_f: zoom difference (focal length difference between two cameras), $\hat{t}_y$: relative shift in Y-direction, $\hat{t}_z$: relative shift in Z-direction, α_x: relative tilt, α_y: relative pan, α_z: relative roll, and $\hat{f}_l$: focal length of left camera in pixels.

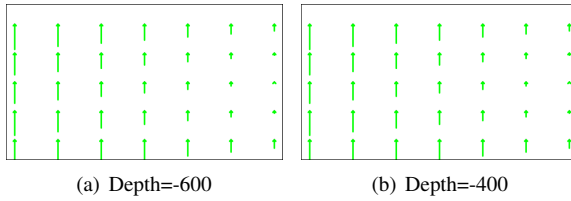(a) Depth=-600   (b) Depth=-400

Figure 23: Vertical disparity at different depths $(u_r - u_l)$ when there are relative tilt and roll angles between the cameras $\alpha_x = 0.1$ and $\alpha_z = 0.1$ respectively.

erty of rotational errors), (2) the value of vertical parallax increases or decreases from the left edge to the right edge of the image (property of roll error), and (3) the value of vertical parallax increases or decreases from the image center to the upper and to the lower edges of the image (property of tilt error). These vertical disparity patterns can be seen in Figure 23.

4.2 Error identification

If there are multiple error sources present simultaneously, it is difficult to identify them from the classified patterns using the proposed classification based approach (shown in Figure 3). Moreover, there can be false matches that create false disparity at certain locations in the map. To overcome these challenges we propose a mathematical approach that solves the unknowns in equation 8. As there are 7 unknowns, namely $\alpha_f, \hat{t}_y, \hat{t}_z/\hat{f}_l, \alpha_x \hat{f}_l, \alpha_x/\hat{f}_l, \alpha_y/\hat{f}_l$ and α_z, at least 7 pairs of $(\mathbf{p}_l, \mathbf{p}_r)$ are needed to solve for the unknowns. To increase the reliability of the solution in the presence of false feature matches, RANSAC algorithm [Fischler81] is used. In this work, RANSAC uses a threshold on Sampson distance to identify false matches. After filtering the false matches, the vertical disparity contributions of each error source are computed for each true match. The error sources that highly influence the total vertical disparity are identified and are considered as the major rig-alignment errors. Figure 22 presents the block diagram of the approach.

5 RESULTS

This section presents the experimental results of the proposed approach. According to my knowledge, this is the first work in this direction and there exists no other work to compare the performance. Consider a stereo image pair (shown in Figure 24) captured using an uncalibrated stereo rig i.e. unknown intrinsic and extrinsic parameters of the cameras. Sparse feature matches are computed using SIFT, SURF and Harris features. Figure 25 show the results of stereo matching. It can be observed that there exists false matches. After applying the RANSAC on equation 8, the false matches are filtered and the unknowns $(\alpha_f, \hat{t}_y, \hat{t}_z/\hat{f}_l, \alpha_x \hat{f}_l, \alpha_x/\hat{f}_l, \alpha_y/\hat{f}_l, \alpha_z)$ are calculated. Figure 26 shows the results of stereo matching after removing false matches. Figure 27 shows the vertical disparity at the true matched features drawn on the left image. It can be observed that the disparity increases from right to left and in the entire image the disparity is upwards. Based on the error to vertical disparity pattern classification, the pattern in the results is close to the one caused by tilt and roll (see Figure 21). To confirm this, a graph that shows the individual contributions of the possible error sources is drawn and the errors that highly influence the total vertical disparity are identified. Figure 28 shows the contributions of each possible error source. It can be observed that tilt and roll are the major alignment errors influencing the total vertical disparity.

(a) Left image   (b) Right image
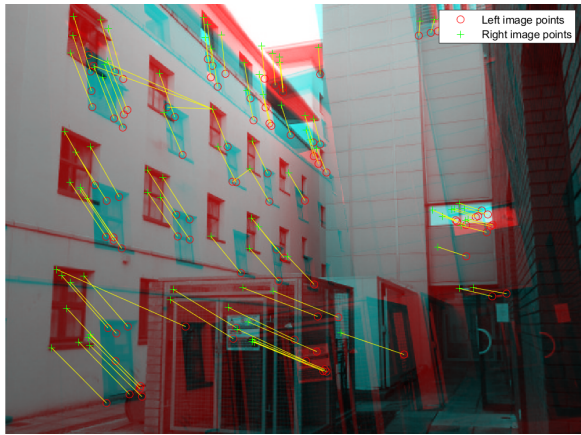
Figure 24: Input stereo image pair.

Figure 25: Visualisation of feature matches (using SIFT, SURF and Harris features) on the right image by drawing a line from the left image point to the right image point.

Figure 26: Feature matches after filtering false matches using RANSAC. The matched points are visualised on the right image by drawing a line from the left image point to the right image point.

6   CONCLUSION

Depending on the type of rig alignment error the vertical disparity is different in different regions of the stereo images. This paper presents a classification of possible disparity patterns to make stereographers or quality analysts identify the error source through visual inspection of the patterns in the vertical disparity map. SIFT, SURF and Harris feature detection algorithms are used to produce the map. This paper also proposed a mathematical approach that computes the contribution of each possible error to the vertical disparity. The advantage is that the approaches do not require any calibration information about the stereo rig. The possible future works include automatic adjustment of rig alignment based on the computed vertical disparity contributions and evaluation of the results using controlled experiments on a real stereo rig setup.

Figure 27: Vertical disparity map drawn on the left image. The red circles indicate the points on the left image and the yellow lines show the vertical disparity at the points.
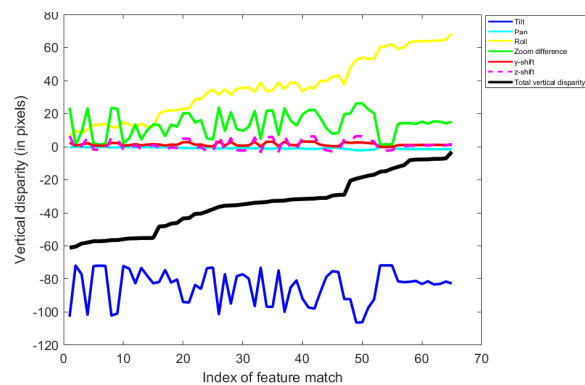

Figure 28: Contributions of possible error sources to the vertical disparity at each feature match. Total vertical disparity is also drawn. The feature matches are sorted in the ascending order of their vertical disparity. Tilt and roll contributions are highly influencing the total vertical disparity.

7   REFERENCES

[Andrea00] Andrea, F., and Emanuele, T., and Alessandro, V. A Compact Algorithm for Rectification of Stereo Pairs, ACM Machine Vision Applications, Vol.12, No.1, pp.16-22, July, 2000.

[ATSC11] Advanced Television Systems Committee (ATSC), Final Report of the ATSC Planning Team on 3D-TV, August, 2011.

[Bay08] Bay, H., and Tuytelaars, T., and Gool, L.V. SURF:Speeded-Up Robust Features, Computer Vision and Image Understanding (CVIU), Vol.110, pp.346-359, June, 2008.

[Boufama95] Boufama, B., and Mohr, R. Epipole and Fundamental Matrix Estimation using Virtual Parallax, Proceedings of the Fifth International Conference on Computer Vision (ICCV), pp.1030-1036, 1995.

[Clarke98] Clarke, T.A., and Wang, X., and Fryer, J.G. The Principal Point and CCD Cameras, The Photogrammetric Record, Vol.16, pp.293-312, Octo-

ber, 1998.

[Faugeras93] Faugeras, O. Three-Dimensional Computer Vision (Artificial Intelligence), The MIT Press, November, 1993.

[Fischler81] Fischler, Martin A., and Bolles, Robert C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography, Communications of the ACM, Vol.24, No.6, pp.381-395, June, 1981.

[Hartley04] , Hartley, R., and Andrew, Z. Multiple View Geometry in Computer Vision, Cambridge University Press, ed.2, April, 2004.

[Hartley92] Hartley, R.I. Estimation of Relative Camera Positions for Uncalibrated Cameras, Proceedings of the European Conference on Computer Vision (ECCV), pp.579-587, 1992.

[Hodges91] Hodges, L.F. Basic Principles of Stereographic Software Development, Proceedings of the SPIE (Society of Photo-Optical Instrumentation Engineers) on Stereoscopic Displays and Applications II, Vol.1457, pp.9-17, 1991.

[IJsselsteijn00] IJsselsteijn, W. A., and de Ridder, H., and Vliegen, J., Effects of stereoscopic filming parameters and display duration on the subjective assessment of eye strain, Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems VII, vol.3957, pp.12-22, May, 2000.

[Kamgar-Parsi88] Kamgar-Parsi, B., and Eastman, R.D. Calibration of a Stereo System with Small Relative Angles, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol.51, No.1, pp.1-19, 1988.

[Knorr12] Knorr, S., and Ide, K., and Kunter, M., and Sikora, T. The Avoidance of Visual Discomfort and Basic Rules for Producing "Good 3D" Pictures, SMPTE Motion Imaging Journal, Vol.121, issue 7, pp.72-79, October, 2012.

[Lenz88] Lenz, R.K., and Tsai, R.Y. Techniques for Calibration of the Scale Factor and Image centre for High Accuracy 3-D Machine Vision Metrology, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.10, pp.713-720, 1988.

[Lowe04] Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints, International Journal on Computer Vision, Vol.60, No.2, pp.91-110, November, 2004.

[Perek16] Perek, P., and Makowski, D., and Napieralski, A. Efficient uncalibrated rectification method for stereo vision systems, MIXDES - 23rd International Conference Mixed Design of Integrated Circuits and Systems, pp. 89-92, June, 2016.

[Sammons92] Sammons, E. The World Of 3-D Movies, Delphi Publication, 1992.

[Tam11] Tam, W.J., and Speranza, F., and Yano, S., and Shimono, K., and Ono, H. Stereoscopic 3D-TV: Visual Comfort, IEEE Transactions on Broadcasting, Vol.57, pp.335-346, 2011.

[Woods93] Woods, A., and Docherty, T., and Koch, R. Image Distortions in Stereoscopic Video Systems. Proceedings of SPIE, Stereoscopic Displays and Applications IV, Vol.1915, pp.36-48, September, 1993.

[Zilly10] Zilly, F., and Muller, M., and Eisert, P., and Kauff, P. Joint Estimation of Epipolar Geometry and Rectification Parameters using Point Correspondences for Stereoscopic TV Sequences, 3D Data Processing, Visualization and Transmission (3DPVT) Conference, May, 2010.

[Zilly11] Zilly, F., and Kluger, J., and Kauff, P. Production Rules for Stereo Acquisition, Proceedings of the IEEE, Special Issue on 3D Media and Displays, Vol.99, issue 4, pp.590-606, April, 2011.