The Stereoscopic Analyzer – An Image-Based Assistance Tool for Stereo Shooting and 3D Production

Frederik Zilly, Marcus Müller, Peter Eisert, Peter Kauff

Fraunhofer Institute for Telecommunications - Heinrich-Hertz-Institut, Berlin, Germany

ABSTRACT

The paper discusses an assistance system for stereo shooting and 3D production, called Stereoscopic Analyzer (STAN). A feature-based scene analysis estimates in real-time the relative pose of the two cameras in order to allow optimal camera alignment and lens settings directly at the set. It automatically eliminates undesired vertical disparities and geometrical distortions through image rectification. In addition, it detects the position of near- and far objects in the scene to derive the optimal inter-axial distance (stereo baseline), and gives a framing alert in case of stereoscopic window violation. Against this background the paper describes the system architecture, explains the theoretical background and discusses future developments.

Index Terms: 3D Production, Stereo Shooting, Image-Based Assistance System, Real-Time 3D Video Analysis

1. INTRODUCTION

It is well known from the past that improper creation of stereo content can easily result in a bad user experience. In fact, the depth impression from a 3D display is a fake of the human visual system and if not done properly, consequences for the human 3D perception might be eye strain and visual fatigue [1]. Production of good stereo content is therefore a difficult art that requires a variety of technical, psychological, and creative skills and has to consider perception and display capabilities.

Therefore, to create good stereo, stereographers have to take into account a variety of conditions, guidelines and rules right from the beginning of the production chain. One main issue is to ensure that the whole scene usually remains within a so-called Comfortable Viewing Range (CVR) of the targeted 3D viewing conditions (e.g. ratio of screen width and viewing distance, also called presence factor). The 3D experience is generally comfortable if all scene elements stay in this limited depth space close to the screen. As the available depth volume is restricted compared to the real 3D world, the difficult job of a stereographer is "*to bring the whole real world inside this virtual space called the comfort zone"* [2].

There are two main parameters by which this production rule can be controlled. One is the inter-axial distance (stereo baseline) which controls the overall range of depth, i.e., the depth volume of the reproduced scene. The other one is the convergence which controls the depth position of the scene in relation to the screen, i.e., which parts of the scene appear behind and which in front of the screen, respectively.

Further issues are the avoidance of undesired effects causing retinal rivalry. This refers to any kind of geometrical distortions (keystones, vertical misalignment, lens distortions, deviations in focal length, etc.), to unbalanced photometry (color mismatches, differences in sharpness, brightness, contrast or gamma, etc.) and to perception conflicts (stereo framing, stereoscopic window violation, extreme out-screening, etc.).

Apart from a mismatching stereo baseline, these deficiencies can usually be corrected in certain limits during post-production. Nevertheless, any careful planning and execution of stereo shooting tries to avoid them from the beginning. This includes an accurate rigging and calibration of the stereo cameras, good adjustment and matching of electronic and optical camera parameters and, above all, the adaptation of the stereo baseline to the depth structure of the scene content. Basically, this adjustment is time-consuming manual work and requires skilled staff to do it properly.

These efforts were accepted as long as 3D productions have only addressed a small niche market. However, due of the rapid increase of 3D productions during the last few years, there is now a rising demand on efficient 3D production tools assisting stereographers and camera team at the set. The main goals of such assistance systems for stereo shooting are to ease rigging, to save time for adjustments, to change them quickly from take to take and to allow also less experienced camera staff to employ proper stereo settings.

Against this background, this paper describes an assistance system for stereo shooting called the Stereoscopic Analyzer (STAN). The next section gives a system overview and describes the general system architecture. Then, section 3 explains the theoretical background of the 3D video analysis used. Finally, section 4 gives an outlook on future developments.

2. SYSTEM DESCRIPTION

The block diagram in **Fig. 1** describes the system architecture and illustrates the signal flow of the STAN. The stereo camera signals are captured using a grabber board with two single-link HD-SDI interfaces.

In a first step of 3D video analysis, the luminance images are down-sampled and a feature detector is used to find interest points and match point correspondences between the two stereo images. The constraints of epipolar geometry are used to identify robust matches, to estimate the pose of the two cameras, and to compute the parameters for the correction of camera misalignments and keystone distortions by rectification. In addition, photometric parameters are analyzed to detect related mismatches and to calculate parameters for matching color, contrast and brightness.

The geometric and photometric correction parameters can either be stored as metadata for later post-production purposes or can directly be used for real-time corrections in case of live broadcasting and for steering the lens control, the electronic camera settings and camera positioning in case of motorized lenses and rigs as well as interfacing to camera signal processing.

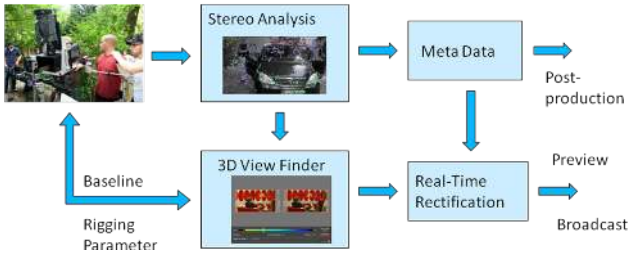

Fig. 1 *System architecture and signal flow of the STAN*

Another main feature of STAN is the 3D viewfinder that provides the stereographer, camera assistant and production staff with information on the stereo geometry and the current camera settings. Related analysis results are visualized at a touch-screen through an intuitive graphical user interface (GUI). Fig. 2 shows some examples.

Fig. 2: *GUI functionalities of STAN's 3D viewfinder*

A special function of the STAN is to analyze the depth structure of the scene and to propose adjustments for basic stereo parameters like baseline and convergence, accordingly. For this purpose a histogram is calculated from the horizontal disparities of the matched point correspondences to detect the near and far objects in the scene (see **Fig. 3**). Although the feature point detection of STAN is already robust, feature point locations might be affected by small errors. Therefore near and far clipping planes are defined as the 2nd and 99th percentile, respectively. The overall disparity range is then the difference between the near- and the far clipping plan.

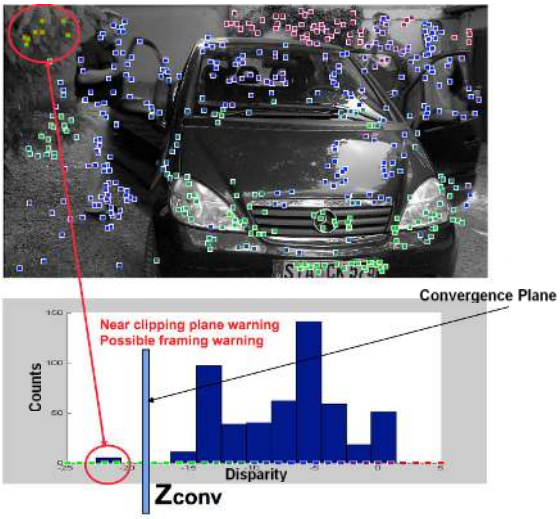

Fig. 3 *Visualization of disparity range and convergence*

As mentioned in the introduction, one main task of stereographers is to fit this disparity range of a given scene into the CVR taking full advantage of the available depth budget of given viewing conditions[3][4]. Once the depth structure of the scene is known, the focal length of the cameras is fixed and the targeted viewing conditions are chosen, the only degree of freedom left is the inter-axial distance. STAN exploits this relation for calculating the optimal inter-axial distance in function of the estimated disparity range, the CVR depth budget and the current inter-axial distance. In the simplest case, it is the "1/30"-rule saying that the CVR behind the screen should not be larger than 1/30 of the screen width [5]. However, any other framework of more complex production rules can be defined and used by the STAN on demand.

Furthermore, the convergence plane Z_{conv} can be shifted interactively within the given depth volume for preview purposes. A color-coded visualization of the feature points indicates out-screening in dependence on Z_{conv} selection and gives frame alert in case of stereoscopic window violation. In parallel, the stereo images are rectified and shifted in real time accordingly to camera pose estimation and Z_{conv} selection, respectively, and can directly be pre-viewed at a 3D control monitor.

3. ANALYSIS OF STEREO GEOMETRY

3.1. Detection of Feature Point Correspondences

The core of STAN is the robust detection of feature point correspondences between the two stereo images. Any suitable feature detectors like SIFT or SURF can be used for this purpose [6][7]. As even these very distinctive descriptors will produce a certain amount of outliers, the search of robust point correspondences is constrained by the epipolar equation from eq. (1). As known from literature, a pair of corresponding points **m** and **m´** in the two stereo images have to respect the epipolar constraint, where **F** denotes the fundamental matrix defined by a set of geometrical parameters like orientations, relative positions, focal lengths and principal points of the two stereo cameras:

$$\mathbf{m}'^{T}\mathbf{F}\mathbf{m} = 0 \tag{1}$$

Based on this epipolar constraint, RANSAC estimation of the fundamental matrix **F** is used to eliminate outliers of feature point correspondences [8]. **Fig. 4** shows an example of related results for images of a stereo test shooting. Note that the cameras are not perfectly aligned in this case and the point correspondences still contain undesired vertical disparities.

Fig. 4: *Robust feature point correspondences for original images of test sequences BEER GARDEN kindly provided by the European FP7 project 3D4YOU*

3.2. Linearization of Epipolar Constraint

It is well-known that the estimation of **F** is numerically challenging. In the STAN application, however, it can be assumed that the stereo cameras have already been mounted in an almost parallel set-up, i.e., the cameras have almost the same orientation perpendicular to the stereo baseline. This means that the camera geometry is already close to the rectified state where **F** degenerates to the following simple relation:

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2}$$

Hence, we can linearize **F** by developing a Taylor expansion around the rectified state from eq. (2) and by ignoring terms higher than first order. In addition, it can be assumed that the principal points are located in the centers of the image sensors, that the difference between the two focal lengths f and f' is small ($f/f'=1+\alpha_f$ with $\alpha_f \ll 1$), that the stereo baseline is defined by the x-axis of the left stereo camera and that the deviations c_y and c_z of the right stereo camera in y- and z-direction across the baseline is small compared to the inter-axial distance c_x along the baseline ($c_y \ll 1$ and $c_z \ll 1$ in case of a normalized baseline $c_x=1$). Under these preconditions, the linearization results in the following simplified term of the matrix **F** where α_x, α_y and α_z denote the orientation angles of the right camera:

$$\mathbf{F} = \begin{bmatrix} 0 & \frac{-\hat{c}_z+\alpha_y}{f} & \hat{c}_y+\alpha_z \\ \frac{\hat{c}_z}{f} & \frac{-\alpha_x}{f} & -1+\alpha_f \\ -\hat{c}_y & 1 & -f\alpha_x \end{bmatrix} \tag{3}$$

Note that the above preconditions are generally fulfilled in case of a proper stereo set-up using professional rigs and prime lenses. Based on this linearization, the epipolar equation from eq. (1) can also be written as follows:

$$\underbrace{v'-v}_{\text{vert. disparity}} = \underbrace{\hat{c}_y\Delta u}_{\text{y-shift}} + \underbrace{\alpha_z u'}_{\text{roll}} + \underbrace{\alpha_f v'}_{\Delta\text{-zoom}} \underbrace{-f\alpha_x}_{\text{tilt-offset in pel.}} \underbrace{+\alpha_y\frac{u'v}{f}}_{\alpha_y\text{-keystone}} \underbrace{-\alpha_x\frac{vv'}{f}}_{\text{tilt ind. keystone}} \underbrace{+\hat{c}_z\frac{uv'-u'v}{f}}_{\text{z-parallax deformation}} \tag{4}$$

This relation can be used to build up a system of linear equations enabling a robust estimation of **F** by RANSAC and to remove outliers from feature point correspondences as described in section 3.1. Furthermore, once **F** has been estimated, its coefficients from eq. (3) can be used to steer and correct geometrical and optical settings in case of motorized rig and lenses.

3.3. Image Rectification

A perfect control of geometrical and optical settings will not be possible in any case. Some stereo rigs are not motorized and adjustments have to be done manually with limited mechanical accuracy. When changing the focus, the focal length of lenses might be affected. In addition, lenses are exchanged during shootings and, if zoom lenses are used, motors do not synchronize exactly and lens control suffers from backlash hysteresis.

As a consequence, slight geometrical distortions may remain in the stereo images. These remaining distortions can be corrected electronically by means of image rectification. The process of image rectification is well-known from literature [9][10][11]. It describes 2D warping functions that are applied to left and right stereo images, respectively, to compensate deviations from the ideal case of parallel stereo geometry. In the particular case, the 2D warping functions are derived from a set of constraints that have to be defined by the given application scenario.

One major constraint in any image rectification is that multiplying corresponding image points **m** and **m´** in eq. (1) with the searched 2D warping matrices **H** and **H´** has to end up with a new fundamental matrix that is equal to rectified state in eq. (2). Cleary, this is not enough to calculate all 16 coefficients of the two matrices **H** and **H´** and, hence, further constraints have to be defined for the particular application case.

Fig. 5: Results of image rectification for the images of test sequence BEER GARDEN from **Fig. 4**

One further constraint in the given application scenario is that the horizontal shifts of the images have to respect the user-defined selection of convergence parameter Z_{conv}. Furthermore, the 2D warping matrix **H** for the left image has to be chosen such that the horizontal and vertical deviations c_y and c_z of the right camera are eliminated, i.e., the left camera has to be rotated such that the new baseline after rectification goes through the focal point of the right camera:

$$\mathbf{H} = \begin{bmatrix} 1 & c_y & fc_z \\ -c_y & 1 & 0 \\ -c_z/f & 0 & 1 \end{bmatrix} \qquad (5)$$

Based on this determination, the 2D warping matrix **H´** for the right image can be calculated straightforwardly by taking into account the additional side-constraints that left and right camera have the same orientation after rectification, that both cameras have the same focal length and that the x-axis of the right cameras has the same orientation as the new baseline:

$$\mathbf{H'} = \begin{bmatrix} 1-\alpha_f & \alpha_z + c_y & 0 \\ -(\alpha_z + c_y) & 1-\alpha_f & -f\alpha_x \\ \frac{\alpha_y - c_z}{f} & -\frac{\alpha_x}{f} & 1 \end{bmatrix} \qquad (6)$$

Fig. 5 shows results from an application of image rectification to the non-rectified originals from **Fig. 4**. Note that all vertical disparities have been eliminated now. A quantitative analysis of the presented rectification algorithm has already been published in [12] and has shown that its performance is similar state-of-the-art rectification methods or even outperforms some of them.

4. CONCLUSIONS AND OUTLOOK

Due to a successful collaboration between Fraunhofer Heinrich Hertz Institute (HHI), Berlin, und KUK Film Production, Munich, long experiences in video analysis and stereo production could be combined and exploited for developing the STAN. **Fig. 6** shows an application of STAN to a mirror rig with two ARRI D21 cameras and a side-by-side rig with two MicroHD cameras from Fraunhofer IIS, as presented for the first time at NAB 2009. A redesigned version will be prototyped for NAB 2010 and will be launched as a product soon. The prototype has already been tested under real working conditions in context with a couple of 3D productions in 2010.

Fig. 6: STAN as presented at NAB 2009

5. REFERENCES

[1] A. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. *Proc. SPIE*, 1915:36–48, Feb. 1993.

[2] B. Mendiburu, "3D Movie Making – Stereoscopic Digital Cinema from Script to Screen." *Elsevier*, 2008.

[3] G. Jones, D. Lee, N. Holliman, and D. Ezra. Controlling perceived depth in stereoscopic images. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems VIII,* Vol. 4297, pages 42-53, June 2001.

[4] G. Sun and N. Holliman. Evaluating methods for controlling depth perception in stereoscopic cinematography. In *Proc. SPIE Stereoscopic Displays and Virtual Reality Systems XX*, Vol. 7237, Jan. 2009.

[5] G. Herbig. The Three Golden Rules of Stereography (in German). S*tereo journal,* Vol. 65, March 2002.

[6] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, Vol. 110, No. 3, pp 346–359, 2008.

[8] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. *Cambridge University Press*, ISBN: 0521540518, second edition, 2004.

[9] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000.

[10] J. Mallon and P. F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23(7):643–650, July 2005.

[11] H.-H.Wu and Y.-H. Yu. Projective rectification with reduced geometric distortion for stereo vision and stereoscopic video. *Journal of Intelligent and Robotic Systems*, 42:71-94(24), 2005.

[12] F. Zilly, M. Müller, P. Eisert, and P. Kauff. Joint Estimation of Epipolar Geometry and Rectification Parameters using Point Correspondences for Stereoscopic TV Sequences. Proceedings of *3DPVT*, Paris, France, May 2010.