Joint Estimation of Epipolar Geometry and Rectification Parameters using Point Correspondences for Stereoscopic TV Sequences

Frederik Zilly, Marcus Müller, Peter Eisert, Peter Kauff

Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institut
Einsteinufer 37, 10587 Berlin, Germany
{frederik.zilly,marcus.mueller,peter.eisert,peter.kauff}@hhi.fraunhofer.de

Abstract

An optimal stereo sequence needs to be rectified in order to avoid vertical disparities and similar image distortions. However, due to imperfect stereo rigs, vertical disparities occur mainly due to a mechanical misalignment of the cameras. Several rectification methods are known, most of them are based on a strong calibration. However, calibration data is often not provided, such that the rectification needs to be done based on point correspondences. In this paper, we propose a rectification technique which estimates the fundamental matrix jointly with the appropriate rectification parameters. The algorithm is designed for narrow baseline stereo rigs. The optical axes are almost parallel (beside a possible convergence angle). The rectification parameters allow a pose estimation of one camera relative to the other one so that the mechanical alignment of the stereo rig can be improved.

1. Introduction

When producing content for 3D cinema or 3DTV, one goal is a perfectly aligned pair of stereo sequences. In fact, any misalignment of the cameras leads to vertical disparities. These vertical disparities in stereo pairs lead to eye strain and visual fatigue [12]. Every stereo rig contains parts of finite mechanical accuracy. Moreover, thermal dilation changes the extrinsic parameters. When changing the lens' focus, the internal parameters are affected, possibly the focal length. In addition, lenses are changed during shootings, and the setup time is limited. When using zoom lenses, the principal point shifts, the focal length is changed over a wide range of values. The motors for zoom level and focus do not synchronize exactly in the general case, so that slightly different focal lengths will occur. Finally, these motors suffer from backlash which can be thought as a hysteresis curve which affects the zoom level. As a consequence, a rectification algorithm is needed which performs reliably and which uses only point correspondences. The resulting image pair should be suitable for watching. Therefore, the rectification method needs to minimize a possible distortion impact. In addition, the convergence plane should not be changed because this is a critical stereo parameter for 3D sensation. As a result, the rectification is not done with respect to the plane at infinity [3] but to a scene dependent plane. The proposed method establishes a relationship between the components of the fundamental matrix, and a physical model of the camera positions. This allows to calculate rectifying homographies with a very small distortion impact. The model assumes a geometry which is near the rectified state such that the fundamental matrix can be expressed as linearized Taylor expansion around the rectified state.

2. Rectification

Image rectification is well known in literature. Faugeras describes rectification as a reprojection of the left and the right image onto a common image plane $\mathcal{R}$ [2]. By rectification he aims to insure a *simple* epipolar geometry, i.e. the epipoles are at infinity and the epipolar lines match with the image scan lines which facilitates dense stereo matching. The new image plane $\mathcal{R}$ needs to parallel to the baseline. However, there are two degrees of freedom to choose such a plane. While one of the degree of freedom only affects a possible scaling of the images, the other parameter is responsible for image distortion effects. Faugeras proposes to choose the plane $\mathcal{R}$ such that it is parallel to the line of intersection of the two original image planes. Zhang et al. propose an algorithm which combines the estimation of the epipolar geometry with a guided point matching [14]. Papadimitriou and Dennis describe a vertical registration algorithm [11]. Hartley proposes a rectification algorithm which is suitable for wide baseline systems [5]. A main idea of this algorithm is to minimize horizontal disparities in order

to facilitate image matching. Furthermore Hartley claims that the image center undergoes a rigid transformation, i.e. only rotation and translation are applied to the image center. Loop and Zhang propose a rectification algorithm which reduces image distortions by decomposing the rectifying homographies into one a similarity transform, followed by a shearing transform [8]. Isgrò and Trucco propose to calculate rectifying homographies without explicit knowledge of the epipolar geometry [7]. Fusiello et al. propose a linear rectification algorithm based on two perspective projection matrices [4]. Wu and Yu propose to minimize distortion by using a properly chosen shearing transform [13]. Mallon and Whelan compare different rectification algorithms with respect to their image distortion impact [10]. They derive rectifying homographies from a fundamental matrix which might be affected by noise.

The aim of all rectification processes is to find two homographies H and H' which, applied to two projection matrices P and P' result in two new matrices P_{rect} and P'_{rect} which have the following properties: Both image planes are parallel, both epipoles are mapped to infinity $(1, 0, 0)$. The result is then that epipolar lines are parallel and coincide with the image scan lines.

2.1. Fundamental Matrix

To establish the point correspondences, which are used for the rectification process, a robust feature detector is needed which produces as few outliers as possible. Suitable Feature detectors are SIFT [9] combined with Difference of Gaussian interest point detection and Up-Right-SURF [1] and the Hessian Box-Filter detector. However, even these very distinctive descriptors will produce a certain amount of outliers. One well known technique is to eliminate outliers using a RANSAC estimation of the fundamental matrix [6]. We will develop a constrained fundamental matrix. The matrix is composed by a set of meaningful parameters (angles, offsets in pixel, difference in focal length). These same parameters will later be used to compose the rectifying homographies. In addition the parameters can be used to optimize the mechanical alignment of the stereo rig.

Given a set of point correspondences m and m' from the left and right camera, they need to fulfill the following equation:

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

with $m = (u, v, 1)$ and $m' = (u', v', 1)$. The initial projection matrices have the form $P = K[I|0]$ and $P' = K'[R|t]$. Following this approach we can describe the initial fundamental matrix F as

$$\mathbf{F} = \mathbf{K'}^{-T} [\mathbf{t}]_\times \mathbf{R} \mathbf{K}^{-1} \tag{2}$$

with

$$\mathbf{t} = -\mathbf{R}\mathbf{C} \tag{3}$$

where $\mathbf{C}$ is the camera center of the right camera. Note that the two inner matrices form the Essential matrix E as in [2]. In the rectified state, we have $\mathbf{K'} = \mathbf{K}$, $\mathbf{R} = \mathbf{I}$ and $\mathbf{t} = (1, 0, 0)^T$. It is easy to show that in this case, the fundamental matrix has the form:

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix} \tag{4}$$

3. Method

Our aim is to develop a Taylor expansion of the fundamental matrix around the rectified state. In order to linearize the algorithm, we cut the Taylor expansion after the first term. The translation vector $\mathbf{t} = (t_x, t_y, t_z)$ can be calculated up to a scale. In our approach we divide $\mathbf{t}$ by t_x which does not affect (1) and denote $\hat{\mathbf{t}} = (1, \hat{t}_y, \hat{t}_z)^T$. As we assume our camera setup to be near the rectified state, we can conclude that $\hat{t}_y \ll 1$ and $\hat{t}_z \ll 1$. We get the following matrix for the translation vector:

$$[\hat{\mathbf{t}}]_\times = \begin{bmatrix} 0 & -\hat{t}_z & \hat{t}_y \\ \hat{t}_z & 0 & -1 \\ -\hat{t}_y & 1 & 0 \end{bmatrix} \tag{5}$$

We assume that the rotation angles are small, and that any second order effects can be neglected ($\alpha \ll 1$). This gives us the following approximation for the rotation matrix:

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

We multiply $[\hat{\mathbf{t}}]_\times$ by $\hat{\mathbf{R}}$ and eliminate any mixed term ($\alpha_x \hat{t}_y = 0, ...$) as second order effect. The resulting Essential matrix E is:

$$\mathbf{E} = \begin{bmatrix} 0 & -\hat{t}_z & \hat{t}_y \\ \hat{t}_z + \alpha_y & -\alpha_x & -1 \\ -\hat{t}_y + \alpha_z & 1 & -\alpha_x \end{bmatrix} \tag{7}$$

Concerning the calibration matrices we assume that the principal point is centered, that the aspect ratio is 1 and that we have zero skew. The focal lengths f and f' can differ. We want to calculate their ratio with high accuracy because this is important in order to avoid vertical disparities induced by a Δ-Zoom. We assume that the ratio $f'/f = 1 + \alpha_f$ where $\alpha_f \ll 1$. We put the origin in the image center. We get the following matrices $\mathbf{K}^{-1}$ and $\mathbf{K'}^{-T}$ respectively where we approximated $1/(1+\alpha_f)$ with $1 - \alpha_f$.

$$\mathbf{K}^{-1} = \begin{bmatrix} 1/f & 0 & 0 \\ 0 & 1/f & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{8}$$

$$\mathbf{K'}^{-T} = \begin{bmatrix} \frac{1-\alpha_f}{f} & 0 & 0 \\ 0 & \frac{1-\alpha_f}{f} & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{9}$$

We can now calculate the linearized fundamental matrix F where we eliminated second order effects ($\alpha_f t_y = 0, \alpha_f \alpha_x = 0 ...$) and multiplied by f:

$$\mathbf{F} = \begin{bmatrix} 0 & \frac{-\hat{t}_z}{f} & \hat{t}_y \\ \frac{\hat{t}_z + \alpha_y}{f} & \frac{-\alpha_x}{f} & -1 + \alpha_f \\ -\hat{t}_y + \alpha_z & 1 & -f\alpha_x \end{bmatrix} \tag{10}$$

We have developed the fundamental matrix with respect to equation (2). We want now substitute t according to (3) and get $\hat{t}_y = \hat{c}_y + \alpha_z$ and $\hat{t}_z = \hat{c}_z - \alpha_y$.

$$\mathbf{F} = \begin{bmatrix} 0 & \frac{-\hat{c}_z + \alpha_y}{f} & \hat{c}_y + \alpha_z \\ \frac{\hat{c}_z}{f} & \frac{-\alpha_x}{f} & -1 + \alpha_f \\ -\hat{c}_y & 1 & -f\alpha_x \end{bmatrix} \tag{11}$$

We insert F into (1)

$$\begin{aligned} & u'\left(v\frac{-\hat{c}_z + \alpha_y}{f} + \hat{c}_y + \alpha_z\right) \\ +v'&\left(u(\frac{\hat{c}_z}{f}) + v\frac{-\alpha_x}{f} - 1 + \alpha_f\right) \\ & + (u(-\hat{c}_y) + v - f\alpha_x) \quad = 0 \end{aligned}$$

We regroup the equation by terms inducing vertical disparities $v' - v$ and have (with $u' - u = \Delta u$).

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

3.1. Estimating the Fundamental matrix

We are now able to build up a system of linear equations which enables us to calculate the coefficients which we need to compose the fundamental matrix. The complete system has the following form:

$$\begin{aligned} \mathbf{Ax} &= \mathbf{b} \\ \mathbf{x} &= (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}. \end{aligned}$$

The vector $\mathbf{b}$ contains the vertical disparities between $\mathbf{m'}$ and $\mathbf{m}$ which are minimized. The result vector $\mathbf{x}$ contains the coefficients which can be used to compose the fundamental matrix and the rectifying homographies:

$$\mathbf{b} = \mathbf{m_2}' - \mathbf{m_2} \tag{12}$$

$$\mathbf{x} = (-f\alpha_x, \alpha_z, \alpha_f, \hat{c}_y, \frac{\alpha_y}{f}, -\frac{\alpha_x}{f}, \frac{c_z}{f})^T \tag{13}$$

$\mathbf{A}$ consists of i rows $\mathbf{A_i}$:

$$\mathbf{A_i} = \begin{pmatrix} 1, u, v', u' - u, u'v, vv', uv' - u'v \end{pmatrix} \tag{14}$$

with $\mathbf{m^i} = (u, v, 1)^T$ and $\mathbf{m'^i} = (u', v', 1)^T$

3.1.1 Model fitting with RANSAC

As one can see, the estimation of c_z depends on four coordinates which makes this estimation numerically unstable. Furthermore, f can be deduced by the two tilt coefficients. When no tilt is present, then this estimation is numerically unstable. In order to use RANSAC, it might be a good choice to omit the estimation of c_z and possibly the estimation of $-\frac{\alpha_x}{f}$. The latter parameter might be neglected when the vertical opening angle is small (as for long shots). Indeed this reduces the number of point correspondences for one guess (sample size) from 7 to 5 (without $-\frac{\alpha_x}{f}$). Furthermore, any pre-knowledge can be exploited. If one knows that $\alpha_f = 0$, this coefficient can be omitted as well. With the same argument, one might omit the estimation of the toe-in α_y if, for instance, the image pair is already dekeystoned. This linearized approach gives us fine granular control of the estimation performance and allows us an insight of the sources of numerical unstableness. The sample size plays an important role in the number of needed samples for the RANSAC, especially when the percentage of outliers is high.

The minimum number of samples is

$$N = log(1 - p)/log(1 - (1 - \epsilon)^s). \tag{15}$$

The following table illustrates this [6] for different sample size and an assumed proportion of outliers $\epsilon = 50\%$ and $p = 99.9$ %. We require a high probability p because we want to do this rectification step a great number of times within a stereoscopic image sequence.

sample size	3	4	5	6	7
required samples	52	108	218	439	881

Table 1. Required samples for RANSAC

Finally, the used distance function for fitting the F-matrix is the Sampson distance:

$$\sum_i \frac{(m_i'^T F m_i)^2}{(Fm_i)_1^2 + (Fm_i)_2^2 + (F^T m_i')_1^2 + (F^T m_i')_2^2} \tag{16}$$

3.1.2 Singularity constraint

The fundamental matrix has rank 2 and hence the determinant should be zero. If our assumption of vanishing second order effects is correct (which of course is the case only up to a certain precision), then the equation ((11)) should lead to a vanishing determinant. However, the numerical value will be non-zero and can be interpreted as an indicator, how well our model of the linearization works. The singularity constraint will not be enforced using the SVD method as described in [6]. In fact, the direct relationship between the components of the fundamental matrix and their physical interpretation as described by (11) would be lost when enforcing the singularity constraint.

3.2. Rectifying Homographies

Once we have roll, tilt, y-shift and Δ-Zoom we can calculate the rectifying homographies directly. Roll and tilt and convergence angle can be corrected by rotating P' in the inverse direction. c_y can be corrected by rotating both cameras around the z-axis (in the same direction by an amount c_y). c_z can be corrected by rotating both cameras around the y-axis (in the same direction by an amount c_z). The offset in Zoom-Level of P' can be corrected with the following homography:

$$\mathbf{H}'_{\mathbf{\Delta-Zoom}} = \begin{bmatrix} 1-\alpha_f & 0 & 0 \\ 0 & 1-\alpha_f & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{17}$$

The rectifying homographies have the form:

$$\mathbf{H} = \mathbf{K}\mathbf{R}^T\mathbf{K}^{-1} \tag{18}$$

$$\mathbf{H} = \begin{bmatrix} 1 & c_y & fc_z \\ -c_y & 1 & 0 \\ -c_z/f & 0 & 1 \end{bmatrix} \tag{19}$$

$$\mathbf{H}' = \begin{bmatrix} 1-\alpha_f & \alpha_z+c_y & 0 \\ -(\alpha_z+c_y) & 1-\alpha_f & -f\alpha_x \\ \frac{\alpha_y-c_z}{f} & -\frac{\alpha_x}{f} & 1 \end{bmatrix} \tag{20}$$

In order to do a rectification with respect to the plane at infinity, the upper-right entry of H' should be $-f(\alpha_y - c_z)$. However, this entry results only in an horizontal offset, which is not wanted in our case. We prefer to preserve the convergence plane.

4. Results

In a first experiment, we applied our method to the dataset supplied by Mallon and Whelan [10]. [1]. Six image pairs were rectified using the point correspondences provided within the dataset. The point correspondences were inserted into the system of linear equations according to 14. To ensure that the results can be compared with [10], all point correspondences were used, without a prior RANSAC filtering. Subsequently, the result vector x defined in (13) which contains the coefficients describing the epipolar geometry was computed. The best performance was achieved when fitting for the following six coefficients: y-shift, roll, Δ-zoom, tilt-offset, α_y-keystone, and tilt-induced keystone. These coefficients were used to build the homographies H and H' according to (19) and (20). The resulting homographies were used to rectify the image pairs. The original and the rectified image pairs are shown in figure 1. In order to perform a quantitative comparision of rectification results, a set of error metric parameters were computed following [10]. For each homography, measures for orthogonality E_o (ideally 90), aspect ratio E_a (ideally 1.0), and rectification error E_r were computed. The values obtained with the proposed method are shown together with the data provided by [10] in table 4. The results regarding Mallon's, Loop's, and Hartley's method were transfered from [10].

Sample	Method	E_o		E_a		Error E_r	
		H'	H	H'	H	Mean	std
Arch	Proposed	89.96	90.00	0.9988	1.0000	0.14	0.36
	Mallon	91.22	90.26	1.0175	1.0045	0.22	0.33
	Loop	95.40	98.94	1.0991	1.1662	131.3	20.63
	Hartley	100.74	93.05	1.2077	1.0546	39.21	13.85
Drive	Proposed	89.98	90.00	0.9992	1.0000	0.01	0.93
	Mallon	90.44	90.12	1.0060	1.0021	0.18	0.91
	Loop	98.73	101.42	1.1541	1.2052	10.41	3.24
	Hartley	107.66	90.87	1.3491	1.015	3.57	3.43
Boxes	Proposed	90.02	90.00	1.0000	1.0000	0.18	0.52
	Mallon	88.78	89.33	0.9785	0.9889	0.44	0.33
	Loop	97.77	95.69	1.1279	1.0900	4.35	9.20
	Hartley	86.56	94.99	0.9412	1.0846	33.36	8.65
Roof	Proposed	90.01	90.00	1.0009	1.0000	0.06	1.15
	Mallon	88.35	88.23	1.1077	0.9700	1.96	2.95
	Loop	69.28	87.70	0.6665	1.0497	0.84	11.01
	Hartley	122.77	80.89	1.5256	0.8552	11.89	18.15
Slates	Proposed	90.00	90.00	1.0001	1.0000	0.23	0.20
	Mallon	89.12	89.13	0.9852	0.9855	0.59	0.56
	Loop	37.29	37.15	0.2698	0.2805	1.14	3.84
	Hartley	89.96	88.54	1.0000	0.9769	2.27	5.18
Yard	Proposed	90.05	90.00	1.0024	1.0000	0.12	0.44
	Mallon	89.91	90.26	0.9987	1.0045	0.53	0.54
	Loop	133.62	134.27	2.1477	2.4045	8.91	13.19
	Hartley	101.95	91.91	1.2303	1.0335	48.19	11.49

Table 4 shows that for our method, the image distortions measured by E_o and E_a are considerably smaller than for any other method. The homography H has always orthogonality $E_o = 90$ and aspect ratio $E_a = 1.0$. As we did not fit for c_z, H results in a 2D rotation around the image center, which does not induce any shearing or anisotropic scaling. The values for H' indicate a very low image distortion. Concerning the rectification error E_r, our method shows good alignment performance. The mean of E_r is nearer to 0 for every image pair. The standard deviation shows an accuracy similar to Mallon's method.

[1] available from http://elm.eeng.dcu.ie/ vsl/vsgcode.html

4.1. Rectification including F-matrix estimation

In a second experiment, we selected one frame of the *Beergarden* stereo sequence. We used SIFT to find putative matches [9]. Subsequently, we used RANSAC to eliminate outliers. To perform one RANSAC iteration, we used 4 point correspondences to fill the constraints matrix A (14) and to fit the results vector x (13) including y-shift, roll, Δ-zoom and tilt-offset . Afterwards, these values were used to compute the candidate fundamental matrix F according to (12). Figure 2 shows the inlying matches for the left and the right image before and after rectification. In figure 3 the original images and the rectified images are overlayed which allows qualitive evaluation of the rectfication performance. The room divider in the background allows for a good inspection of the alignment of the two cameras. Apparently, the rectification process resultet in a well aligned image pair.

5. Conclusion

We have proposed a rectification technique for which allows to compute rectifying homographies as well as the computation of a fundamental matrix. The latter is important to use the algorithm during a RANSAC elimination of outliers. The algorithm uses point correspondences and does not need prior knowlegde of the projection matrices. The technique involves a linearized computation of the epipolar geometry which makes it suitable for setups which are near the rectfied state. The image distortions have shown to be neglectable compared to techniques proposed in [8], [5], and [10]. The algorithm preserves the convergence plane of the stereo setup and is suitable for rectifying 3DTV stereo sequences [12].

References

[1] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer Vision ECCV 2006*, Lecture Notes in Computer Science, chapter 32, pages 404–417. 2006. 2

[2] O. Faugeras. *Three-Dimensional Computer Vision (Artificial Intelligence)*. The MIT Press, November 1993. 1, 2

[3] A. Fusiello and L. Irsara. Quasi-euclidean uncalibrated epipolar rectification. In *ICPR08*, pages 1–4, 2008. 1

[4] A. Fusiello, E. Trucco, and A. Verri. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22, 2000. 2

[5] R. I. Hartley. Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127, November 1999. 1, 5

[6] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition, 2004. 2, 3, 4

[7] F. Isgrò and E. Trucco. Projective rectification without epipolar geometry. In *CVPR '99*, pages 1094–1099. IEEE Computer Society, 1999. 2

[8] C. Loop and Z. Zhang. Computing rectifying homographies for stereo vision. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, page 131 Vol. 1, 1999. 2, 5

[9] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, November 2004. 2, 5

[10] J. Mallon and P. F. Whelan. Projective rectification from the fundamental matrix. *Image and Vision Computing*, 23(7):643–650, July 2005. 2, 4, 5

[11] D. Papadimitriou and T. Dennis. Epipolar line estimation and rectification for stereo image pairs. *Image Processing, IEEE Transactions on*, 5(4):672–676, Apr 1996. 1

[12] A. Woods, T. Docherty, and R. Koch. Image distortions in stereoscopic video systems. *Proc. SPIE*, 1915:36–48, February 1993. 1, 5

[13] 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), January 2005. 2

[14] Z. Zhang, R. Deriche, O. D. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1995. 1

(a) Arch

(b) Drive

(c) Boxes

(d) Roof

(e) Slates

(f) Yard

Figure 1. From left to right: original left, original right, rectified left, rectified right

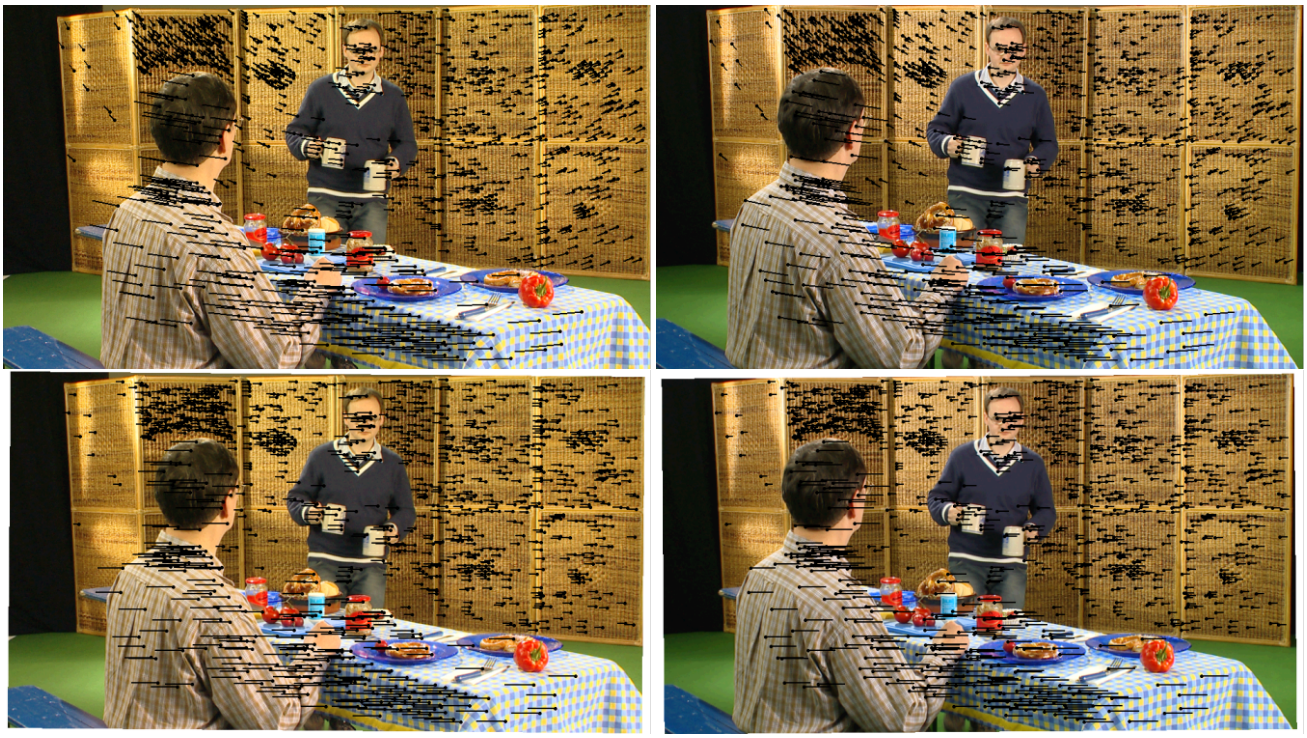

Figure 2. From left to right and top to bottom: original left, original right, rectified left, rectified right

Figure 3. A stereo pair from the Beergarden Sequence. Overlay of the two original images (top) and the rectified images (bottom).