Visual-Geometric Scene Reconstruction from Image Streams

Reinhard Koch and Jan-Michael Frahm

Multimedia Information Processing Group
Institute of Computer Science and Applied Mathematics
Christian-Albrechts-University of Kiel
Herman-Rodewald-Str. 3, 24098 Kiel, Germany
Email: rk@informatik.uni-kiel.de

Abstract

Visual-geometric scene reconstructions capture the visual appearance and geometry of real 3D scenes from image streams. Depending on the type of camera motion and type of scene structure, either 2D od 3D representations can be used. From the representations one can render new views of the scene with high degree of realism. Recently we see a lot of research in this field, especially for the case that uncalibrated handheld camera are utilized as input devices. We will give an overview of the different approaches and representations.

1 Introduction

Reconstruction of 3D scene models from image streams only is in general an ill-posed problem [5]. Therefore one must apply certain constraints on the recording conditions or on the scene in order to obtain reliable reconstructions. Therefore we will need to develop representations that allow to reduce modeling complexity and methods for model selection are needed. In this contribution we will discuss some of the imaging situations that occur in practice and outline the neccessary algorithms for modeling and representing scenes.

Certain constraints on camera motion and scene structure allow to select specific representation models. In case of a purely rotating camera or with 2D background scenes only one may obtain a 2D mosaic scene representation. This representation type is modelled by global scene-to-image homographies that can be measured robustly from the image sequence. When full 3D camera motion and full 3D structure is present one must resort to a 3D structure from motion approach that is able to completely model the surfaces of 3D scene objects from a projective reconstruction based on local image correspondences. Between these two extreme representations one may choose from a variety of 2D-3D representations that model the scene with a varying amount of depth information. Examples are image based rendering methods like image interpolation or the lumigraph.

We will discuss these representations and their application to an uncalibrated recording situation, meaning that neither the camera motion nor the intrinsic camera parameters are controlled or calibrated beforehand. We will develop a framework that handles the continuum of 2D- to 3D-representations, based on a selective switching of the underlying representation model. Self-calibration of the intrinsic camera parameters is included in the approach.

2 Single viewpoint imaging

Visual-geometric 3D scene reconstruction has been investigated thoroughly over the last decades. We can distinguish different types of analysis, based on the type of camera motion and scene structure. In the case of purely rotating cameras with fixed focal point, no 3D scene structure can be observed and we may obtain a single viewpoint panorama.

In the case of a camera that rotates around its projection center, we obtain a single-viewpoint panoramic image by stitching together different images into a panoramic image representation. All images are viewed from the same single focal point but with different viewing directions. The camera is modeled as pinhole projection $P_{4x3} = KR[I|-C]$ with an affine intrinsic parameter matrix K and extrinsic orientations R, C. I is the identity matrix, K holds the camera parameters focal length f, aspect ratio a, skew s and image center point $[c_x, c_y]^T$:

$$K = \begin{bmatrix} f & s & c_x \\ 0 & af & c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

R is the orthonormal camera rotation matrix and C the camera projection center vector, both fixed relative to an arbitrary world coordinate system. Given a homogeneous 3D scene point $M = [X, Y, Z, 1]^T$ and a set of cameras with projection matrices P_j, the scene point M is projected onto the corresponding non-homogeneous image points $m_j = [x_j, y_j, w_j]^T$ of the cameras.

$$m_j = K_j R_j [I | - C] M \quad , \quad (2)$$

Since we have a fixed projection center C for all cameras, we may without loss of generality choose $C = 0$ which yields

$$m_j = K_j R_j [I|0] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = K_j R_j \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

The projection depends only on K_j and R_j, not on C. Since the last coordinate of M is irrelevant, one may also obtain the same result by projecting the direction vector $\vec{M} = [X, Y, Z]$ to the camera. $\vec{M}$ can be eliminated from eq. (3) if it is observed in two images i and k at corresponding image points m_i and m_k:

$$\begin{aligned} m_i &= K_i R_i \vec{M} \Rightarrow \vec{M} = R_i^{-1} K_i^{-1} m_i \\ m_k &= K_k R_k \vec{M} = K_k R_k R_i^{-1} K_i^{-1} m_i. \end{aligned} \quad (4)$$

We can see that a projective 2D transformation H_{ik} exists between the corresponding image points m_i and m_k:

$$m_k = H_{ik} m_i \ \text{ with } \ H_{ik} = K_k R_k R_i^{-1} K_i^{-1}. \quad (5)$$

If at least four corresponding points [1] between an image pair can be found we can compute H_{ik}.

Once H_{ik} is available we can transfer all image points from image i to k and vice versa. This observation forms the basis for many algorithms to align images in panoramas [16, 25, 33]. Overlapping images of a rotating camera are taken and form

[1] We need eight linearly independent equations to solve for the eight independent parameters of H_{ik}. Thus the four corresponding point coordinates must form a quadrilateral in the image with not more than two points on the same straight line. In that case the matrix is of rank 3 and its inverse exist.
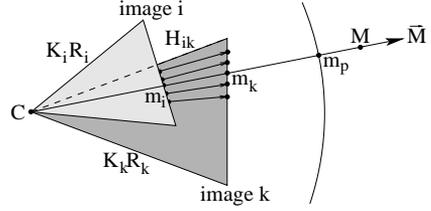

Figure 1: Registration of rotated cameras with planar the homography H_{ik} and projection of the corresponding image points m_i, m_k along the ray $\vec{M}$ onto the panoramic surface at m_p.

an *image mosaic* that is stitched together onto a new image with larger field of view. The new image need not be planar but may be warped around the camera focal point to obtain full 360° field of view. Depending on the dominant rotation axis one may use zylindrical or spherical coordinates for the panoramic image. This imaging situation is shown in fig. 1 where point correspondences $m_k = H_{ik} m_i$ can be mapped between the images and onto the panoramic imaging surface at m_p.

2.1 Self-calibration from camera rotation

The homograpy H explains the mapping between overlapping images but does not distinguish between camera intrinsics and rotation. While R determines the global orientation of the camera coordinate axes, K determines the viewing direction of projection rays for each pixel w.r.t. the global orientation. If we want to map the images onto a global panoramic map, say a sphere, we need to know K. This can be established by camera self calibration. Let us assume that we can compute H_{ik} from image point correspondences. The relative rotation between both images is denoted by $R_{ik} = R_k R_i^{-1}$ and can be expressed as $R_{ik} = K_k^{-1} H_{ik} K_i$ from eq. (5). Since R is an orthonormal rotation matrix, $R = R^{-T}$, hence

$$\begin{aligned} R_{ik} &= K_k^{-1} H_{ik} K_i = K_k^T H_{ik}^{-T} K_i^{-T} \\ \Rightarrow (K_k K_k^T) &= H_{ik} (K_i K_i^T) H_{ik}^T \end{aligned} \quad (6)$$

Equation (6) relates the Calibration Matrices K_i, K_k to the Homography H_{ik} and can be used for self-calibration of K from the image point correspondences. If the calibration matrix K is equal for both images we can directly estimate it from H_{ik}

since enough equations are available. In the case of varying K additional contraints need to be imposed, like keeping some parameters constant or by using prior knowledge. One reasonable assumption for example is to set the image skew s to zero.

The approach was investigated in detail by Hartley [12] for constant K and recently by Agapito et. al. [1] for changing intrinsic camera parameters. More information about panoramic imaging can be found in [3].

3 Multiple viewpoint imaging

The single viewpoint panorama does not contain any depth information since we considered the projection of the viewing rays only. If the camera center is moved between images then we must consider the full projection model described in eq. (2), despicted in fig. 2. Since M is projected onto different projection centers C_j for each camera, the projection rays for the corresponding points m_j differ and can not be canceled out like in eq. (6). Instead we obtain a depth-dependent parallax between image correspondences. If we consider corresponding image points only, the projection ray from one image point m_i projects onto a line (1-parameter family of possible correspondences m_k) in the other image. Hence we can not have a point-point-correspondence as in the case of single-viewpoint homographies but rather a point-line transfer of possible correspondences, the epipolar constraint. The point-line-transfer is described by the Fundamental matrix F_{ik} between image pairs. Given the camera projection centers C_i, C_k of the camera pair and the image point m_i, these three points form an epipolar plane that intersects both image planes at the epipolar lines l_i, l_k. The epipolar constraint states that a

Figure 2: Epipolar geometry between two images.

corresponding point m_i must be on the corresponding epipolar line l_i induced by m_k:

$$m_i^T l_i = 0 \text{ with } l_i = F_{ki} m_k \qquad (7)$$

Substitution yields the Fundamental Matrix constraint

$$m_i^T F_{ki} m_k = 0. \qquad (8)$$

F_{ik} can be seen as a generalization of H_{ik} for the case that both camera centers differ. A derivation of F can be found in many textbooks, see [5, 27]. In fact, $F_{ik} = [e]_x H_{ik}$ with $[e]_x$ the skew-symmetric cross-product operator of the vector e. e is the intersection of the line $C_i - C_k$ with the image plane and is called the epipole. F_{ik} can be computed if three conditions hold (see [27]):

1. at least seven independent point correspondences are available,
2. not all corresponding 3D points are allowed to lie on a plane, and
3. the two camera projection centers must differ (otherwise we can not compute e).

3.1 Self-calibration from multiple viewpoints

The Fundamental Matrix is the only constraint that can be derived from image correspondences of an uncalibrated camera alone. Since it is independent of scene depth it can be used to restrict the search range for image correspondence to a one-dimensional line ambiguity. Corresponding point may then be found by local correspondence search along the epipolar line. Robust methods [34] exist for this approach. Unfortunately, one can not recover the full projection matrices P_i, P_k directly from F_{ik} since F_{ik} gives only seven independent measurements. The full projection matrix on the otherhand has 11 d.o.f. since we have five intrinsic calibration and six extrinsic orientation parameters. Like was the case with calibration of varying K from H_{ik} we need more information. Therefore we need to evaluate multiple F_{ik} from sequences of image pairs. In addition some of the calibration parameters must be set constant or known, as described before in section 2.1. A wealth on literature about self-calibration has emerged over the last years. Starting with pioneering work by Hartley [11] and Faugeras [6] the problem has been investigated by numerous authors [14, 29, 8].
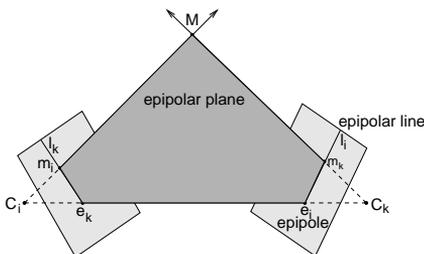
4 The full 3-D case: Structure from Motion

With all these information at hand we are able to devise a strategy for full 3-D reconstruction from image sequences. If a moving camera observes images of a rigid scene we can recover 3D scene structure and camera parameter from image correspondences alone.

This approach is termed *structure from motion* and is described in detail in [30]. We will summarize it in the following section.

Beardsley et al. [2] proposed a scheme to obtain projective calibration and 3D structure by robustly tracking salient feature points throughout an image sequence. This sparse object representation outlines the object shape, but gives not sufficient surface detail for visual reconstruction. Highly realistic 3D surface models need the dense depth estimation and can not rely on few feature points alone.

In [29] the method of Beardsley e.a. was extended in two directions. On the one hand the projective reconstruction was updated to metric even for varying internal camera parameters, on the other hand a dense stereo matching technique [4] was applied between two selected images of the sequence to obtain a dense depth map for a single viewpoint. From this depth map a triangular surface wire-frame was constructed and texture mapping from one image was applied to obtain realistic surface models. In [19] the approach was further extended to multi viewpoint sequence analysis.

We employ a 3-step approach for reconstruction:

- camera self-calibration and metric structure is obtained by robust tracking of salient feature points over the image sequence,
- local depth maps for each viewpoint are computed from correspondences between adjacent image pairs of the sequence,
- all correspondence maps are linked together by multiple view point linking to fuse depth measurements over the sequence.

Two-view analysis: The basic tool for the camera tracking is the two-view matcher. Image intensity features are detected with the Harris corner detector [10] and have to be matched between the two images I_i, I_k of the view points P_i, P_k. We rely on a robust computation of the Fundamental matrix F_{ik} with the RANSAC (RANdom SAmpling Consensus) method [34]. A minimum set of 7 seed features correspondences is picked from a large list of potential image matches to compute a specific F_{ik}. For this particular F the support is computed from the other potential matches. This procedure is repeated with random seeds to obtain the most likely F_{ik} with best support in feature correspondence.

The next step after establishment of F_{ik} is the computation of the camera projection matrices P_i and P_k. The fundamental matrix alone does not suffice to fully compute the projection matrices. In a bootstrap step for the first two images we follow the approach by Beardsley *et al.* [2]. Since the camera calibration matrix K is unknown a priori we assume a approximate $\tilde{K}$ to start with. The first camera is then set to $P_0 = \tilde{K}[I|0]$ to coincide with the world coordinate system, and the second camera P_1 can be derived from the epipole e and F as

$$P_0 = \tilde{K}[I|0 , \ P_1 = \tilde{K}\left[[e]_x F + ea^T|re\right]$$

P_1 is defined up to a global scale r and the unknown plane $\pi_{\inf}$, encoded in a^T (see also [30]). Thus we can only obtain a projective reconstruction. The vector a^T should be chosen such that the left 3×3 matrix of P_i best approximates an orthonormal rotation matrix. The scale r is set such that the baseline length between the first two cameras is unity. K and a^T will be determined later during camera self-calibration.

Once we have obtained the projection matrices we can triangulate the corresponding image features to obtain the corresponding projective 3D object features. The object points are determined such that their reprojection error in the images is minimized. In addition we compute the point uncertainty covariance to keep track of measurement uncertainties. The 3D object points serve as the *memory* for consistent camera tracking, and it is desirable to track the projection of the 3D points through as many images as possible.

Sequential camera tracking: Each new view of the sequence is used to refine the initial reconstruction and to determine the camera viewpoint. Here we rely on the fact that two adjacent frames of the sequence are taken from nearby view points, hence many object features will be visible in both views. The procedure for adding a new frame is much like the bootstrap phase. Robust matching of $F_{i,i+1}$ between the current and the next frame of the sequence relates the 2D image features between views

I_i and I_{i+1}. Since we have also the 2D/3D relationship between image and object features for view I_i, we can transfer the object features to view I_{i+1} as well. We can therefore think of the 3D features as self-induced calibration pattern and directly solve for the camera projection matrix from the known 2D/3D correspondence in view I_{i+1} with a robust (RANSAC) computation of P_{i+1}. In a last step we update the existing 3D structure by minimizing the resulting feature reprojection error in all images. A Kalman filter is applied for each 3D point and 3D position and covariance are updated accordingly. Unreliable features and outliers are removed, and newly found features are added.

Retrieving the metric framework: The viewpoint mesh contains the projective calibration only. For a metric representation we have to remove the projective skew from the scene using self-calibration methods. The topic of self-calibration is a research field in itself. We refer to the relevant literature on this. Ways to find the projective transformation by applying constraints on the internal camera parameters were discussed in depth by e.g. [36, 29].

4.1 Dense stereo pair matching

With the camera calibration given for all viewpoints of the sequence, we can proceed with methods developed for calibrated structure from motion algorithms. The feature tracking algorithm already delivers a sparse surface model based on distinct feature points. This however is not sufficient to reconstruct geometrically correct and visually pleasing surface models. This task is accomplished by a dense disparity matching that estimates correspondences from the grey level images directly by exploiting additional geometrical contstraints.

Exploiting scene constraints: The epipolar constraint obtained from calibration restricts corresponding image points to lie in the epipolar plane which also cuts a 3D profile out of the surface of the scene objects. The profile projects onto the corresponding epipolar lines in the image pair where it forms an ordered set of neighboring correspondences.

For well behaved surfaces this ordering is preserved and delivers an additional constraint, known as 'ordering constraint'. Scene constraints like this can be applied by making weak assumptions about the object geometry. In many real applications the observed objects will be opaque and composed out of piecewise continuous surfaces. If this restriction holds then additional contraints can be imposed on the correspondence estimation.

The pairwise disparity estimation allows to compute image to image correspondence between adjacent rectified image pairs, and independent depth estimates for each camera viewpoint. An optimal joint estimate will be achieved by fusing all independent estimates into a common 3D model. Finally we can apply surface texture mapping to add visual fidelity to the model.

5 2D/3D Representations

Both single- and multi-viewpoint approaches have advantages and drawbacks. If only stationary rotating cameras are in use one would certainly resort to single-viewpoint representations. This situation is in fact a common case. Typical examples are cameras observing sport events, or fixed-mounted pan-tilt-surveillance cameras. Robust and reliable techniques for panorama generation are available and the representation is of high fidelity. However, in interaction with the representation one is restricted to 2-D operations and no true 3D-interaction (like mixing with 3-D objects) is possible.

Multi-viewpoint representation on the other hand allow (in principle) full 3-D interaction with the scene. However, one is bound to the requirements as described above. Therefore we always need sufficient 3-D camera motion and general scene structure to avoid degenerate situations and singularities. Also, it is much harder to obtain high fidelity 3-D models in unrestricted scenes. In theory we would need reliable 3-D structure estimates for every possible 3-D scene point which might not be feasible. Fortunately, sometimes we do not need a full 3-D scene representation but rather a restricted subset. Such representations may contain either accurate depth from certain view points or for certain areas only [37]. Examples of such partial representations are discussed in the sections below.

Model selection: When dealing with real imaging situations and uncalibrated handheld cameras we can not distinguish the type of scenes and motion beforehand. Rather is it neccessary to *selectively switch* between different modes during image processing. Torr et.al. [35] develloped a model

switching strategy based on simultaneous evaluation of H_{ik} and F_{ik}. This model allows to detect different motion models as well as certain degenerate cases in scene structure like planar scenes. This approach can even be carried further in the case of a rotating camera, since one is able to distinguish certain rotational components from the homograpy and adapt the number of model parameters to avoid parametric overfitting. Depending on the selected model we can then uns one of the following partial scene representations.

Polycentric and stereoscopic panoramas: Polycentric panoramas mix the wide viewpoint capabilities of a panorama with stereoscopic visualisation. Recently a whole new field has emerged here and impressive results were obtained (see [28, 15, 31]. These representation is mainly thought for stereoscopic visualisation since no depth is computed.

Layered depth representations: Layered depth [32] can be thought of as a coarse depth estimation by using occlusion maps and stacked images. We could also think of sets of panoramas that are placed one after the other. It is useful when only restricted 3-D interaction is needed for object sorting with before/behind decisions.

Plane-parallax representations: Plane-parallax representations [23] are used when the scene is composed mostly of a planar component plus some restricted deviation from it. Typical examples are geographical mappings from an airplane onto the earth surface. In this case we can construct a hight map over the dominant plane [17, 18].

Depth maps: Depth maps are the traditional approach to viewpoint-dependent depth, sometimes termed 2.5 D [5, 26]. For each camera view point one may generate a complete depth representation that can be used to interact with the scene. The approach is similar to the plane-parallax model, based on the image plane as reference plane.

5.1 Image interpolation

Depth dependent image interpolation is located in between 2-D and 3-D methods. It tries to combine the fidelity and computational ease of polycentric images with the interactive power of 3-D structure. A special group of such representation is the light field that has received a lot of attention lately in the graphics community. The light field captures the apearence of an object by recording an exhaustive set of sample images from different view points. These images store the reflected appearance (basically the BRDF for a special lighting situation) of the object from all possible directions.

In the basic approach one renders a novel view of the object from the image collection by interpolating incident viewing rays from nearby images [24]. However, the interpolation tends to blur the rendering if there is a depth parallax between the images selected for interpolation. Depth-dependent blurring can be avoided either if very many closely spaced image are taken or when the depth parallax is compensated by image warping as done in the Lumigraph approach [9]. For the warping we need to estimate the depth parallax which is done by computing 3-D structure. However, in contrast to a full 3-D approach that needs a em globally consistent depth estimate, we need only to compute the *local depth map* between adjacent pairs of images. The advantage is

- local depth maps between nearby viewpoints are easier to obtain
- even if no dense depth map can be computed, we can use a coarse (planar) depth approximation in a scalable representation,
- the approach can model viewpoint-dependent reflections since the surface texture is switched depending on the view point.
- global depth inconsistencies do not show since they are *covered* by texture mapping,

The disadvantage of the approach is the high data volume (many images and depth maps have to be stored) and the computational load for the offline-computation of the depth maps.

5.2 The uncalibrated lumigraph

We can combine the Structure-from-Motion Approach (SfM) as discussed in section 4 with view-dependent rendering of a lightfield that was acquired from a handheld camera with uncalibrated imaging [20, 21]. The advantage of this approach is that it combines the strength of SfM (camera tracking and local depth map computation from a freely moving camera) with the rendering qualities of the lightfield (fast rendering, arbitrary reflectance, approximate depth modeling). Thus we obtain a versatile system for rendering arbitrary scenes with high fidelity [22].

In contrast to standard lightfields that require a

highly calibrated setup, the camera can be swept freely over the scene to form a generic surface of camera view points, the viewpoint surface. SfM then tracks the cameras and computes the best calibration by considering all images and their epipolar relations in a two-dimensional camera mesh [20]. For each camera we obtain the corresponding depth map and store them together with the color image. To render from such a generalized light field, we first determine the cameras that project into the novel viewpoint. The image content of these cameras is then rendered into the novel view via a planar homography in two steps. First, we locally approximate the 3D surface by a plane and project the original image onto this plane via a planar homography H_{j1} and in a second step this texture map[2] is then projected into the novel view with a second homography H_{j2}. Multiplication of both matrices yield the final mapping from original to novel view which is implemented efficiently in OpenGL hardware by a projective map [22].

6 Conclusion

In this contribution we have given an overview of different representations for scene reconstruction from uncalibrated image streams. Goal is a highly realistic rendering of the given scene from novel view points. Depending on the type of motion and scene structure one may select 2D panoramic or 3D geometric representations for this task. One must be careful to handle the degeneracies associated with each motion and scene model and should be able to automatically select the appropriate model to avoid parametric overfitting.

We reviewed the full 3D approach for structure from motion computation of a freely moving camera and discussed different intermediate representations that capture partial scene structure. We then showed how to combine different approaches into a consistend framework, which leads to a mixed geometric and image-based representation.

References

[1] L. de Agapito, E. Hayman, E. Reid, "Self-calibration of rotating and zooming cameras", *Int. Journal Comp. Vision, Kluwer, Boston. To appear in 2001.*

[2] P. Beardsley, P. Torr and A. Zisserman: 3D Model Acquisition from Extended Image Sequences. In: B. Buxton, R. Cipolla(Eds.) Computer Vision - ECCV 96, Cambridge, UK., vol.2, pp.683-695. Lecture Notes in Computer Science, Vol. 1064. Springer 1996.

[3] R. Benosman and S.B. Kang (eds.): "Panoramic Vision." *Springer Verlag, 2001.*

[4] L.Falkenhagen: Hierarchical Block-Based Disparity Estimation Considering Neighborhood Constraints. Intern. Workshop on SNHC and 3D Imaging, Rhodes, Greece, Sept. 1997.

[5] O. Faugeras: Three-dimensional computer vision: a geometric approach. MIT-Press, Cambridge, MA, 1993.

[6] O. Faugeras: What can be seen in three dimensions with an uncalibrated stereo rig. *Proc. ECCV'92*, pp.563-578.

[7] O. Faugeras, Q.-T. Luong and S. Maybank: Camera self-calibration - Theory and experiments. *Proc. ECCV'92*, pp.321-334.

[8] A. Fusiello, "Uncalibrated Euclidean reconstruction: a review", *Image and Vision Computing 18 (2000), pp. 555-563, 2000.*

[9] S. Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen: The Lumigraph. Proceedings SIGGRAPH '96, pp 43–54, ACM Press, New York, 1996.

[10] C.G. Harris and J.M. Pike: 3D Positional Integration from Image Sequences. 3rd Alvey Vision Conf, pp. 233-236, 1987.

[11] R. Hartley: Estimation of relative camera positions for uncalibrated cameras. *ECCV'92*, pp.579-587.

[12] R. Hartley, " Self-Calibration of stationary Cameras ", *Int. Journal Comp. Vision 22(1),5-23, Kluwer, Boston.*

[13] B. Heigl, R. Koch, M. Pollefeys, J. Denzler: " Plenoptic Modeling and Rendering from Image Sequences taken by a Hand-Held Camera." *Proceedings DAGM'99, Bonn, Sept. 1999.*

[14] A. Heyden and K. Åström: Euclidean Reconstruction from Image Sequences with Varying and Unknown Focal Length and Principal Point. *Proc. CVPR'97.*

[15] F. Huang,, S.K. Wei, R. Klette, "Epipolar Geometry in Polycentric Panoramas", *in:*

---

[2]This mapping onto the surface is in fact a view-dependent texture mapping, since for each novel view the surface texture is regenerated from the best original camera view.

Klette, Huang, Gimmel'farb (eds.), Multi-Image Analysis, LNCS 2032, pp. 39-50, Springer Verlag, 2001.

[16] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, " Mosaic Representations of Video Sequences and Their Applications." *Signal Processing: Image Communication, special issue on Image and Video Semantics: Processing, Analysis, and Application, Vol. 8, No. 4, May 1996.*

[17] M. Irani, P. Anandan, and D. Weinshall, " From Reference Frames to Reference Planes: Multi-View Parallax Geometry and Applications." European Conference on Computer Vision (ECCV), June 1998.

[18] M. Irani, P. Anandan, and Meir Cohen, "Direct Recovery of Planar-Parallax from Multiple Frames." *ICCV'99 Workshop: Vision Algorithms 99, Corfu, September 1999.*

[19] R. Koch, M. Pollefeys, and L. Van Gool, " Multi Viewpoint Stereo from Uncalibrated Video Sequences." *Proc. ECCV'98, Freiburg, June 1998.*

[20] R. Koch, M. Pollefeys, B. Heigl, L. Van Gool, H. Niemann, " Calibration of Hand-held Camera Sequences for Plenoptic Modeling."*Proc. of ICCV'99, Korfu, Greece, Sept. 1999.*

[21] R. Koch, M. Pollefeys and L. Van Gool: Robust Calibration and 3D Geometric Modeling from Large Collections of Uncalibrated Images. Proceedings DAGM'99, Bonn, Sept. 1999.

[22] R. Koch, B. heigl, "Image-based Rendering from Uncalibrated Lightfields with Scalable Geometry", *in: Klette, Huang, Gimmel'farb (eds.), Multi-Image Analysis, LNCS 2032, pp. 51-67, Springer Verlag, 2001.*

[23] R. Kumar, P. Anandan, K. Hanna, "Direct recovery of shape from multiple views: a parallax based approach", *Proceedings 12th Int. Conf. Pattern Recognition ICPR 94*

[24] M. Levoy, P. Hanrahan: Lightfield Rendering. Proceedings SIGGRAPH '96, pp 31–42, ACM Press, New York, 1996.

[25] S. Mann, R. Picard, "Video Orbits: characterizind the coordinate transformation between two images using the projective group", *Proceedings of Int. Conf. Computer Vision ICCV 95, Boston, MA, USA.*

[26] D. Marr, "Vision", *Freeman, 1982*

[27] T. Moons: A Guided Tour through Multiview Relations. In: Koch, Van Gool (eds.), 3D Structure from Multiple Images of Large Scale Environments. *LNCS Series Vol. 1506, pp. 304-346. Springer-Verlag, 1998.*

[28] S. Peleg, M. Ben-Ezra, "Stereo panorama with a single camera", *Int. Conf. Computer Vision and Pattern Recognition ICPR 99, pp. 395-401, Ft. Collins, Colorado, USA, 1999.*

[29] M. Pollefeys, R. Koch and L. Van Gool: " Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters. " *International Journal of Computer Vision 32 (1), 7-25, 1999.*

[30] M. Pollefeys, R. Koch, M. Vergauwen and L. Van Gool: Metric 3D Surface Reconstruction from Uncalibrated Image Sequences. In: Koch, Van Gool (eds.), 3D Structure from Multiple Images of Large Scale Environments. *LNCS Series Vol. 1506, pp. 139-154. Springer-Verlag, 1998.*

[31] S. Seitz, "The space of all stereo images", *Int. Conf. Computer Vision ICCV01, Vancouver, Can, 2001.*

[32] J. Shade, S. Gortler, L. He, R. Szeliski, "Layered depth images", Proceeding Computer Graphics SIGGRAPH 98.

[33] R. Szeliski and H. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps. " *Computer Graphics (SIGGRAPH'97), pages 251-258, August 1997.*

[34] P.H.S. Torr: Motion Segmentation and Outlier Detection. PhD thesis, University of Oxford, UK, 1995.

[35] P. Torr, A. Fitzgibbon, A. Zisserman, "The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences", *International Journal Computer Vision 32 (1), pp. 27-44, August, 1999.*

[36] B. Triggs: The Absolute Quadric. *Proc. CVPR'97.*

[37] S.K. Wei, F. Huang, R. Klette, "Classification and Characterisation of Image Acquisition for 3D Scene Visualization and Reconstruction Applications", *in: Klette, Huang, Gimmel'farb (eds.), Multi-Image Analysis, LNCS 2032, Springer Verlag, 2001.*