PROJECTIVE RECONSTRUCTION OF BUILDING SHAPE FROM SILHOUETTE IMAGES ACQUIRED FROM UNCALIBRATED CAMERAS

Po-Lun Lai and Alper Yilmaz

Photogrammetric Computer Vision Lab
Ohio State University, Columbus, Ohio, USA
-lai.138@osu.edu, -yilmaz.15@osu.edu
http://dpl.ceegs.ohio-state.edu/

Commission III/2

KEY WORDS: Digital Photogrammetry, Computer Vision, Reconstruction, Slicing, Urban Planning, Visualization

ABSTRACT:

Recovering the three-dimensional (3D) object shape lies as an unresolved and active research topic on the cross-section of computer vision, photogrammetry and bioinformatics. Although various techniques have been developed to tackle the shape recovery problems, the computational complexity and the constraints introduced by the other algorithms have limited the applicability of these methods in real world problems. In this paper, we propose a method that is based on the projective geometry between the object space and silhouette images taken from multiple viewing angles. The approach eliminates the requirements of dense feature matching and camera calibration that are generally adopted by other reconstruction method. The object is reconstructed by setting a set of hypothetical planes slicing the object volume and estimating the projective geometric relations between the images. The experimental results show that satisfactory 3D model can be generated by applying minimal constraints.

1. INTRODUCTION

The growing demands on using 3D models for planning and analysis make 3D object shape recovery a prevailing area of research in the fields of digital photogrammetry and computer vision. Numerous research efforts have been extended to recover the 3D object shape from images. The most commonly adopted methods require calibrating the camera prior to the shape recovery by measuring certain features on a calibrating box which is preset in the object space. The 3D object shape is then reconstructed by traditional triangulation techniques and bundle block adjustment (Blostein and Huang, 1987). Alternatively, the calibration can be accomplished by exploiting the projective geometry which relates the different views of a scene to each other (Hartley et al., 1992), (Koch et al., 2000), and (Hernandez et al., 2007). In general, these methods estimate either the fundamental matrix or the homography between the images and the 3D object shape is recovered up to an unknown scale factor.

While having calibrated cameras is desirable for generating satisfactory result, the calibration process is usually not intuitive. Hence, it becomes necessary to develop techniques to recover the object shape without calibration. Another problem observed is that most 3D object shape reconstruction algorithms depend on the extraction of feature points which poses a limitation in cases when the number of feature points is insufficient or the image is of poor quality. The deformation of the object shape due to the projection from 3D to 2D also increases the difficulty of finding the correspondences between images. In addition, the object occlusion, which is commonly observed during the imaging process, can obstruct the recovery of the object shape.

In the recent years, various papers which are dedicated to the problem of 3D object shape recovery have utilized the properties of the homography transformation and the silhouette images to solve the aforementioned problems. The homography transformation provides a strong geometric constraint and is comparatively simple. The implied 3D scene information can be retrieved from 2D images by the homography transformation and is utilized for use in many applications, including but not limited to tracking people (Khan and Shah, 2006), shadow removal, and detecting low-lying objects (Kelly et al., 2005). These methods are diversified from the planar homography to the infinite homography, and from a single image to multiple images. Although some techniques have conceptually been proven to be successful in certain cases (Zhang and Hanson, 1996) (Wada et al., 2000) (Zhang et al., 2003) (Yun et al., 2006), in real-world problems their use is limited due to specific requirements or assumptions such as positioning the cameras on a circle enclosing the object. The computational complexity also hinders them from being practical.

In this paper, we exploit a new method for the metric reconstruction of 3D object shape using the concept of slicing planes. Our method is inspired by the affine recovery technique proposed in (Khan et al., 2007). However, in our method the constraints used therein are eliminated, and metric shape recovery is obtained. We represent a 3D object by a set of parallel planes intersecting with the object in the object space. These hypothetical planes are related by the homography transformation. Assume four conjugate points lying on a base plane in the object space can be identified in all images, they are sufficient for constructing the homography relation between the images. To generate the view of another hypothetical plane that is parallel to this base plane, we utilize the concept of a vanishing point (Hartley and Zisserman, 2004). The vanishing point of a reference direction is computed from the image of a pair of parallel lines, after that any new set of four points which forms a plane parallel to the base plane can be derived along the reference direction by setting an increment value. For every set of points, the homographies between a reference image and all other images are computed. After warping all images onto the reference image by the homography transformation, their intersection provides the object shape.

The merit of the proposed approach can be described in terms of efficiency, flexibility and practicability. First of all, this method requires no camera calibration or the estimation of the fundamental matrix; hence, it reduces the computational complexity by eliminating the requirement for abundant conjugate points. The object is reconstructed using the contours enclosing the intersected regions, which are sufficient for revealing the complete object surface without the necessity of estimating visual hulls. Second, the formulation provides three different settings for finding the hypothetical planes, using either a pair of parallel lines, or any four randomly located features with known heights in the object space. Furthermore, the level of detail in the reconstructed object is easily modified by changing the number of images used or the density of the planes to find a best balance between the computation time and the smoothness of the recovered 3D object shape. Since no dense point correspondences are needed and the missing information can be recovered from other images of different views, the use of object silhouettes automatically eliminates the problems related to occlusion. Another noteworthy advantage is that we adopt full homography, so that true metric reconstruction is accomplished by the proposed method.

The paper is organized as follows. In order to better manifest the core techniques adopted by our method, we briefly review two important concepts of projective geometry in Section 2.1. The proposed approach for finding consecutive parallel planes is described in Section 2.2. In Section 2.3 the techniques for recovering the 3D object shape using the slices are delineated. Two sets of experiments are conducted to verify the applicability in close-range and aerial photogrammetry. The setup and results are discussed in Section 3. Finally we conclude the paper in Section 4.

2. OBJECT SHAPE RECOVERY

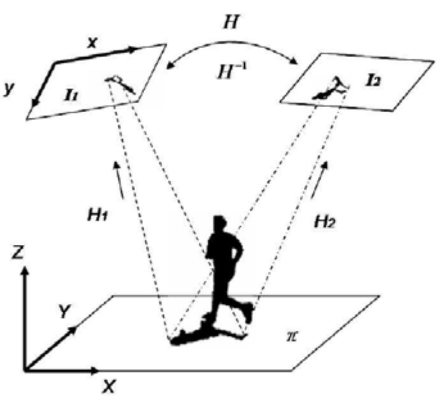

Figure 1: Homography transformation from ground plane to images and across images

The relation between the 3D object space and the 2D image space can be expressed in terms of the projective geometry, which defines a set of invariant properties when object space is projected onto an image space. These invariants in turn provide capabilities to relate two or more views of a scene by projecting one onto the other via mapping equations. In order to provide basic principles of the mapping equations used in this paper, we first provide an introductory discussion to the projective geometry as it pertains to 3D shape recovery which will be detailed later in the section.

2.1 The Projective Geometry

The projective geometry describes the physical characteristics of the cameras and the relationships between the images. The projection of a point $\mathbf{X}_w$ in the object space to a point $\mathbf{x}_i$ in the image space using a projective camera is expressed in terms of a direct linear mapping in homogeneous coordinates as:

$$\lambda \mathbf{x} = \mathrm{P}\mathbf{X}_w = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \mathbf{p}_4] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}, \tag{1}$$

where λ is the scale factor due to projective equivalency of (*kx; ky; k*) = (*x; y;* 1), P is a 3 x 4 camera projection matrix and $\mathbf{p}_i$ the i^{th} column of P. Note that, throughout the paper, we use homogeneous coordinates for points both in the image and object spaces. In the homogenous representation, the last row of the vectors being equal to 1 reflects that the point lies on the image plane. Let's consider the case when the point in the object space lies on the ground plane such that $Z = 0$, then the linear mapping given in (1) will reduce to the planar homography

$$s\mathbf{x} = \mathrm{H}\mathbf{X}_w^{'} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_4] \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \tag{2}$$

where H is the homography matrix which maps points lying on a plane in the object space across different images. This formulation introduces another scaling factor, *s*, to the mapping equation which stems from setting *Z* to 0. One thing to be kept in mind is that the "ground plane" is not necessarily the real ground in object space but can be any visible plane in the scene. Since general 3D modeling rarely requires absolute geodetic coordinates, a local axis system usually fits well in most cases.

In the case when we have multiple images of a scene, an intuitive consequence of the homography transform from the ground plane to the image is the existence of a direct mapping between the two images:

$$\mathbf{x}_i = \mathrm{H}_{wi}\mathbf{X}_w^{'} = \mathrm{H}_{wi}(\mathrm{H}_{wj}^{-1}\mathbf{x}_j) = \mathrm{H}_{ji}\mathbf{x}_j. \tag{3}$$

The H_{ji} matrix is the homography describing the projective transformation between the images *i* and *j*. The estimation of this transformation up to a scale factor requires a minimum of four points lying on the plane. This is exemplified in Figure 1, where the correspondences between ground plane π and images I_1 and I_2 are related by two different homographies. When warping one image onto the other, only the pixels in the area that map the ground plane coincide, while other pixels will create discrepancies depending on their Euclidean distances to the plane π. We should state that it is these discrepancies that will let us estimate the 3D shape of the object.

The perspective effects during the imaging process causes in 3D objects appearing as stretched out into infinity. Considering a pair of parallel lines $\mathbf{L_1}$ and $\mathbf{L_2}$ in the object space, the intersection point is defined to lie at the infinity which is represented by $[X, Y, Z, 0]^T$. When these parallel lines and their intersection at the infinity are imaged, they are mapped into two non-parallel lines and a visible intersection point between them which is referred to as the vanishing point $\mathbf{v}$. The vanishing point can be computed from the cross product of the corresponding line pair $\mathbf{l_1}$, $\mathbf{l_2}$ in the image space (see Figure 3),

$$\mathbf{v} = \mathbf{l}_1 \times \mathbf{l}_2. \tag{4}$$

A vanishing point depends only on the direction of the lines, which means despite their positions all parallel lines with the same directions intersect at one single point. Though parallelism is not preserved after projection, the information about orientation implied by the vanishing point still plays the role of a key to achieve camera calibration and 3D object shape reconstruction.

2.2 Generating the Slicing Planes

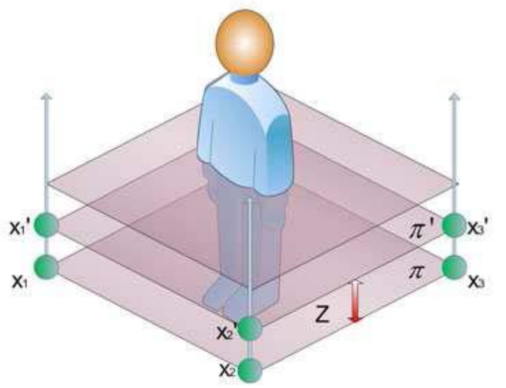

Figure 2: A set of hypothetical planes intersect the object volume and create slices in the object space.

The tools used in our method can be considered as the extensions of the projective geometry concepts delineated in the previous section. In a nutshell, the basic idea behind the proposed approach is to divide the object space into a set of planes parallel to each other as shown in Figure 2. These planes and the homography transform of each of them onto the images generate silhouette coherency maps which provide the 3D shape information.

Let's assume four points $\mathbf{X}_i$, $i = 1, 2, 3, 4$ lie on a pair of parallel lines in π. Imagine that these points are moved up a distance Z vertically in the direction of the plane normal generating four new positions $\mathbf{X}_i^{'}$, $i = 1, 2, 3, 4$. By definition, the new point set constitute a new plane π' which is parallel to π. When the projective camera maps the original four points and the new four points onto the image, the resulting points respectively can be denoted as $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$, $i = 1,2,3,4$ in the image space. In order to automatically estimate the new point set $\mathbf{x}_i^{'}$ directly from the originating planeπwe need to apply an additional constraint. Assume $\mathbf{x}_1^{'}$ and $\mathbf{x}_3^{'}$ can be observed on the image and the height Z is known, then $\mathbf{x}_2^{'}$ and $\mathbf{x}_4^{'}$ can be computed by exploiting the intrinsic properties of the vanishing points. The procedure can be described as follows. First, the vanishing point of the parallel lines in $\pi$is computed using

$$\mathbf{v} = (\mathbf{x}_1 \times \mathbf{x}_2) \times (\mathbf{x}_3 \times \mathbf{x}_4). \tag{5}$$

The vanishing point $\mathbf{v}_z$ in the direction of the normal of plane π is obtained similarly using a pair of lines that is orthogonal to the plane π in the object space. The next step is to establish the relationship between $\mathbf{x}_i$ and $\mathbf{x}_i^{'}$. Rewriting equation (1) as:

$$\lambda_i \mathbf{x}_i^{'} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_4] \begin{bmatrix} X_i \\ Y_i \\ 1 \end{bmatrix} + \mathbf{p}_3 Z. \tag{6}$$

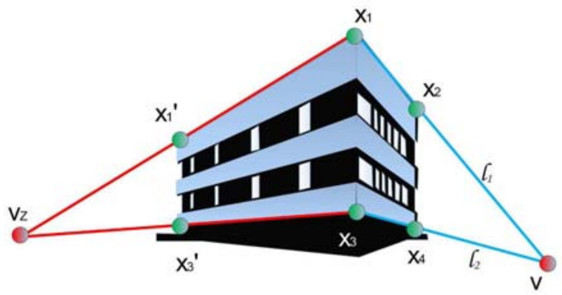

Figure 3: Vanishing points of parallel lines. The points $\mathbf{x}_i$ constitute the reference plane. By finding $\mathbf{x}_1^{'}$, $\mathbf{x}_3^{'}$, $\mathbf{v}$ and $\mathbf{v}_z$, any plane that is parallel to the reference plane can be determined.

The column vector $\mathbf{p}_3$ corresponds to the vanishing point in the direction of the Z axis or the normal of the ground plane. By substituting $\mathbf{p}_3$ with $\mathbf{v}_z$ and combining with equation (2) results in:

$$\lambda_i \mathbf{x}_i^{'} = s_i \mathbf{x}_i + \mathbf{v}_z Z. \tag{7}$$

The unknowns of this linear equation are λ_i and s_i which are described in equations (1) and (2). Estimating both λ_i and s_i can be achieved by solving

$$\begin{bmatrix} \lambda_i \\ s_i \end{bmatrix} = (\mathrm{A}_i^T \mathrm{A}_i)^{-1} \mathrm{A}_i^T \mathbf{b}_i, \tag{8}$$

where $\mathrm{A_i} = [\mathbf{x}_i^{'} \mid -\mathbf{x}_i]$ and $\mathbf{b_i} = \mathbf{v}_z Z$.

Once s_i is computed, estimation of any image point along the line $\mathbf{x}_i\mathbf{v}_z$ is achieved by setting different Z to different set of values. In the case when only x_{o1} and x_{o3} are identifiable on the image, as shown in Figure 3, $\mathbf{x}_2^{'}$ and $\mathbf{x}_4^{'}$ can also be estimated from the property of vanishing point, such that, all parallel lines in the object space which are in the same direction intersect at the vanishing point. If $\mathbf{v}$ is computed from equation (5), then $\mathbf{x}_2^{'}$ and $\mathbf{x}_4^{'}$ are obtained by

$$\begin{aligned}\mathbf{x}_2^{'} &= (\mathbf{x}_2 \times \mathbf{v}_z) \times (\mathbf{x}_1^{'} \times \mathbf{v}) \\ \mathbf{x}_4^{'} &= (\mathbf{x}_4 \times \mathbf{v}_z) \times (\mathbf{x}_3^{'} \times \mathbf{v})\end{aligned}. \tag{9}$$

Assume multiple images of a scene are provided and the first image is chosen as the reference image, such that the other images are warped onto this reference image by $I_{ij} = \mathrm{H}_{ij}I_j$, where the subscript ij indicates that the warping is from i^{th} to j^{th} image. In the sequel of a segmentation method, the object silhouettes extracted in these images highlights of the slicing planes when they intersect with the object volume. Let's consider the object silhouette is defined by setting the image pixels inside the object to 1. The highlights of the slicing planes when they intersect the object volume are determined by warping all the silhouettes onto the reference image:

$$I_{intersection} = \frac{I_1 + \sum_{i=2}^{n} I_{i1}}{n}, \tag{10}$$

where n is the number of images. Next, we will discuss how these warped silhouettes can be used to recover the object shape.

2.3 Recovering the 3D Object Shape

In equation (10), thresholding the accumulation of the warped silhouettes provides a mask image. This mask is the image of the intersection between the slicing planes and the object volume. Hence, using these masks, we can generate the outlines of the object shape which corresponds to the surface of the object volume.

The back-projection of masks generated from equation (10) can be achieved in various ways. Given some feature points with known absolute object coordinates, the relation between the object space and the image space for back-projecting the intersection images onto the object space and create the exact 3D model. If a specific feature such as a box or a building with known dimensions or relative length ratio is recognized in the images, we can assume a local Euclidean coordinate frame in the object space and the metric shape recovery is achieved up to a scale. By selecting an arbitrary local coordinate frame, we can also recover the object shape in the case of distortions on the silhouettes. Theoretically, one can pick up any measurable feature on the reference plane even though the axes are not orthogonal, but ideally a square is preferred for achieving metric reconstruction. On the extreme case the coordinate information of the object space is not available, the 3D object shape of a specific viewing angle can still be represented, as will be shown in the next section.

3. RESULTS AND DISCUSSION

In order to verify the proposed method, we have performed two sets of experiments. In the first experiment, as shown in Figure 4(a), we placed a toy, which contains irregular shape, on the ground plane. The ground plane contains squares which provide us with four measures required to compute the homography transform from one image to the other and from the reference image to the object space. Two pens oriented in the normal direction of the plane are placed to estimate the vanishing point in the direction of Z axis as well as the scale factor s. We took 11 images of the object and the vertical features from different viewpoints around the toy. We should note that, no length measurements are performed and the lengths of the vertical features are set to be a unit length. Since no knowledge of absolute ground truth is considered, we assume all tiles are squares and the coordinates of four corners are defined to reside at unit distances from the origin. The 3D shape is reconstructed by setting the distance increments ΔZ in the vertical direction to 0.5 and computing corresponding Z values used to generate slicing planes. In order to generate fine 3D models, one can set ΔZ to lower values.

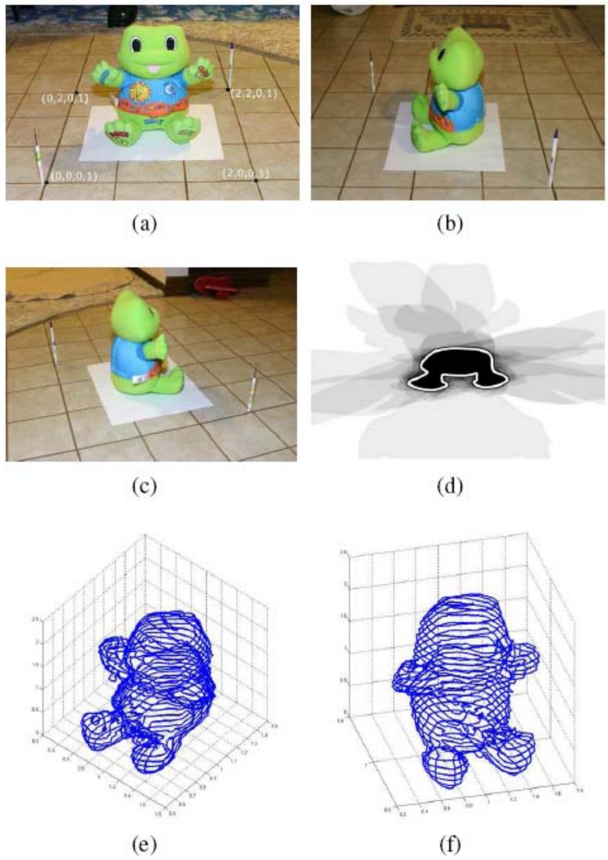

Figure 4: Experiment on recovering the 3D shape of a toy. (a), (b) and (c) show the original images taken from different aspects. The contour of the darkest region in (d) is used for generating 3D points. Two different views of the reconstructed 3D shape are shown in (e) and (f).

The scales and lengths of three axes are not absolute, and as a consequence, the scale of reconstructed object may vary. In order to accelerate the process, only edge pixels of the intersecting silhouettes (Figure 4(d)) are involved in computation. Over 24,000 densely distributed 3D surface points are generated in this test. Among the pictures the toy has several self-occluded parts. The result shows that the effect of occlusion is compensated with information provided by other views and the reconstructed model is intact. Details such as shapes of the toes and hands are able to be observed from the rebuilt model shown in Figure 4(e) and 4(f).

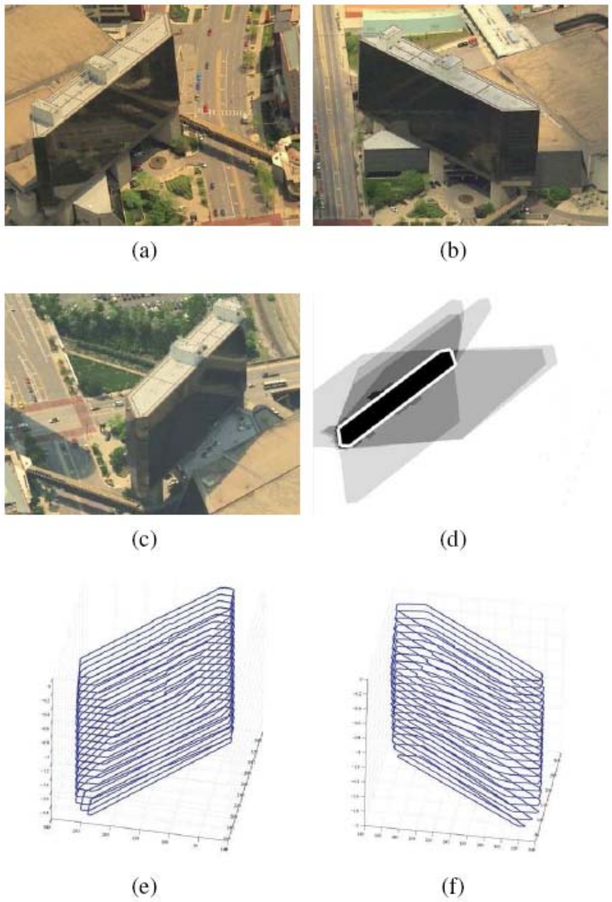

Figure 5: Experiment on recovering the 3D shape of a building. (a), (b) and (c) are the east, north and west views of the building. The brightest region in (d) indicates the slice image on the reference plane. Two different views of the reconstructed 3D building shape are shown in (e) and (f).

The second experiment takes 4 screenshots from the website of http://maps.live.com (Figure 5(a), 5(b) and 5(c)). The "ground plane" is set on the top of a building and 4 conjugate points are measured on each image. The same procedure as in the first experiment is performed except that no object coordinate frame is presumed and the object shape is reconstructed "without back-projection". The object space is set to be identical as the reference image. All the contours of slice images are warped by homography onto this space with preset Z values that are used in computing the hypothetical planes. The result suggests that, in the case when no ground truth is available and true metric reconstruction is not necessary, we can still recover the 3D shape up to a scale factor (Figure 5(e) and 5(f)). This experiment also demonstrates that for an object of a relatively regular shape, images from 4 views are sufficient for achieving satisfactory reconstruction.

4. CONCLUSION

The proposed approach in this paper reconstructs the 3D object shape by exploiting silhouette images taken from uncalibrated cameras. The reconstruction is a metric recovery up to a scale factor which can be determined if object space measurements are provided. The silhouette images are allowed to contain occlusions and distortions as long as some other views of the object reveal the occluded regions. The projective geometric relations between the images provide an easy to implement algorithm. Compared to other algorithms, which require generation of the convex hull or estimation of the fundamental matrix, the proposed approach bears lower computational complexity. The requirement of having abundant feature correspondences in other prevailing techniques is also removed to increase the computational efficiency. The optimized balance between running time and accuracy is determined by the number of images and number of slicing planes. Additional post processing which has not been applied in this paper can be used to further improve resulting 3D surfaces. The experimental results show the applicability of our method for building 3D models from close-range or aerial images. A variety of applications such as urban and rural surface modeling and glacier and polar icecap monitoring can be realized by our method.

REFERENCES

Blostein, S. and Huang, T., 1987. Quantization error in stereo triangulation. In: IEEE Int. Conf. on Computer Vision.

Hartley, R. and Zisserman, A., 2004. *Multiple View Geometry in computer Vision-second edition*. Cambridge Un. Press.

Hartley, R., Gupta, R. and Chang, T., 1992. Stereo from uncalibrated cameras. In: IEEE Conf. on Computer Vision and Pattern Recognition.

Hernandez, C., Schmitt, F. and Cipol, R., 2007. Silhouette coherence for camera calibration under circular motion. In: IEEE Trans. on Pattern Analysis and Machine Intelligence.

Kelly, P., Beardsley, P., Cooke, E., O'Connor, N. and Smeaton, A., 2005. Detecting shadows and low-lying objects in indoor and outdoor scenes using homographies. In: IEE International Conference on Visual Information Engineering.

Khan, S. and Shah, M., 2006. A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: European Conf. on Computer Vision.

Khan, S., Yan, P. and Shah, M., 2007. A homographic framework for the fusion of multi-view silhouettes. In: IEEE Int. Conf. on Computer Vision.

Koch, R., Pollefeys, M. and Gool, L., 2000. Realistic surface reconstruction of 3d scenes from uncalibrated image sequences. In: Visualization and Computer Animation.

Wada, T., Wu, X., Tokai, S. and Matsuyama, T., 2000. Homography based parallel volume intersection: Toward real-time volume reconstruction using active cameras. In: IEEE International Workshop on Computer Architectures for Machine Perception.

Yun, Y., Kim, S., Lee, S., Kim, D. and Choi, J., 2006. Threedimensional reconstruction of an object using three-planar homography of a single image. In: Optical Engineering, Vol. 45.

Zhang, Q., Wang, H. and Wei, S., 2003. A new algorithm for 3d projective reconstruction based on infinite homography. In: Machine Learning and Cybernetics.

Zhang, Z. and Hanson, A., 1996. 3d reconstruction based on homography mapping