



# Novel View Specification and Synthesis

Keith Connor and Ian Reid  
Department of Engineering Science  
University of Oxford  
Parks Road, OX1 3PJ, UK  
[krc, ian]@robots.ox.ac.uk

## Abstract

Given a set of real images, Novel View Synthesis (NVS) aims to produce views of a scene that would correspond to that of a virtual camera. There exist many approaches to solving this problem. We consider physically valid NVS methods, in particular those based on epipolar and trifocal transfer. We review a number of methods and place them into a common framework of view specification and mapping method. We present a new method for directly specifying the novel camera motion for epipolar transfer. We also develop a backward mapping scheme for trifocal transfer which overcomes the problems associated with standard forward mapping methods.

## 1 Introduction

The problem of Novel View Synthesis (NVS) captures the duality between the disciplines of computer graphics and computer vision. Graphics, in a basic sense, seeks to produce images of scene data, whether it be actively acquired real data or synthetic. Vision on the other hand is mostly the reverse; it is the process of obtaining salient data and higher level descriptions from real images.

The standard reconstruction paradigm is dependent upon explicit 3D model building and rendering, i.e. Model-Based-Rendering. In contrast, with Image-Based-Rendering we seek to bypass the model construction and perform direct image-to-image processing. This brings a number of benefits, principally: (i) shorter computation chain, (ii) processing independent of scene complexity.

One possibility is to opt for some form of interpolation scheme, having many similarities to the morphing techniques of computer graphics. Interpolation methods seek to produce *in-between* views of a scene, i.e. views that correspond to the image from a camera located on, or close to, the baseline of the original views. The image-to-image processing here involves interpolating the disparity field to predict the new location for all correspondences [10]. Interpolation is only physically valid under certain assumptions and prerequisites [3]. More recent work not guaranteed to produce physically valid views is that of Lhuillier [5]. They aim to produce interpolation and handle occlusion for general scenes by using an image-based structure they call the *Joint View Triangulation*.

In contrast, methods that exploit constraints from imaging geometry allow a greater range of novel views. These geometry-based methods are often referred to as reprojection



or transfer methods. A key element of transfer is a warping function, i.e. the image-to-image processing consists of a direct mapping from original, corresponding image elements to elements in another view. Work in this area mostly uses matching constraints obtained from epipolar geometry [4] [2].

In this paper we consider the geometry based transfer methods. We make the following contributions:

- A structured review of physically valid transfer methods for NVS
- A common framework for evaluation of transfer methods
- A direct view specification for epipolar transfer
- A backward mapping scheme for trifocal transfer.

## 2 2D Transfer

In 2D, points may be transferred from one image to another by a homography. A homography is the most general linear mapping of 2D projective space. It may be represented by a  $3 \times 3$  matrix,  $H$ , where 2D points in homogeneous coordinates are transformed according to  $p' = Hp$ , with,  $p = (x, y, 1)^T$  and  $p' = (sx', sy', s)^T$ . A homography is used in one of two ways to effect transfer: forward mapping or backward mapping.

### 2.1 Forward vs. Backward Mapping

In the following each point is represented as a homogeneous vector.  $p$  is a point in the source image and  $p'$  a point in the destination image.

**Forward** each pixel of the source image is transformed according to  $p' = Hp$ , then  $p'$  is rounded to the nearest integer pixel location and colour or intensity is copied.

**Backward** each pixel of the destination image is transformed according to  $p = H^{-1}p'$ , then  $p$  is a sub-pixel position at which the colour/intensity to copy may be interpolated (usually bilinear interpolation of neighbouring pixels).

Forward mapping has several problems: because image resolution is finite many pixels may map to one pixel and, even worse, gaps may be left in the resulting image. Thus, forward mapping suffers order dependency and aliasing effects due to integer rounding. In contrast, backward mapping ensures no gaps from finite resolution because it locates the transferred colour for each destination pixel and it can limit aliasing by interpolation. Figure 1 illustrates the properties of each mapping mode and shows a real image warped by a homography. We refer back to these issues when dealing with each of the 3D transfer methods.

## 3 3D Transfer

We now turn to the problem of determining the appearance of a third (novel) view given two or more real views. This section will present epipolar, line-of-sight and trifocal transfer. We outline each algorithm and describe in detail the relationship between them. Finally, we present a novel backward mapping method for trifocal transfer.

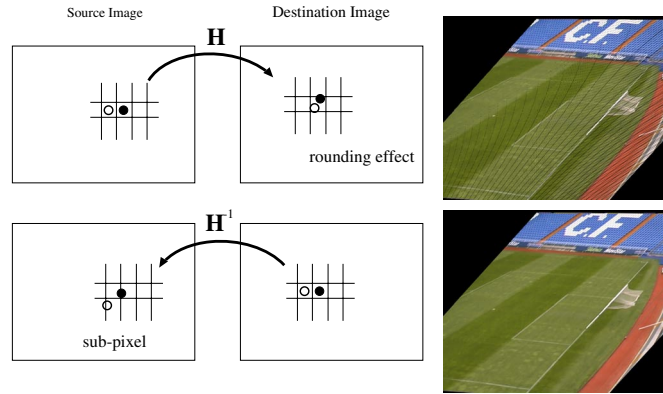


Figure 1: (top) Forward Mapping; (bottom) Backward Mapping. Thin black ‘tracks’ due to finite resolution can be seen in the forward mapped version.

### 3.1 Epipolar Transfer

Faugeras et al. [1] describe how the pairwise epipolar geometry between a third view and each of the first and second views predicts the location and properties of corresponding elements in the third view. For a given element in one image the epipolar constraint reduces the possible location of the correspondence in another view from the whole image to a line in the image. Epipolar transfer exploits the situation where we have two views in correspondence and wish to produce a third view for which we know the fundamental matrices relative to the original images. Figure 2 illustrates the configuration. A correspondence between original images ( $p \leftrightarrow p'$ ) constrains the point in the third image  $p''$  to lie on the lines  $F_{31}p$  and  $F_{32}p'$ . The point in the third view must be the intersection of these two epipolar lines and is given by:

$$p'' = F_{31}p \times F_{32}p'. \quad (1)$$

Assuming dense correspondence between two original views and the required view defined in terms of fundamental matrices, a novel view can be produced by transferring all corresponding pixels to the new view by this method. The transfer is a forward mapping and has the following problems:

1. Forward mapping is not ideal for view synthesis (see Figure 1)
2. Specifying the novel view by identifying (often manually) at least seven point correspondences in the new view to obtain the fundamental matrices is indirect and tedious
3. Degenerate configurations are possible.

Item 1 is dealt with by [2] which describes a backward mapping method. Each pixel in the new view defines an epipolar line in each original view as show in Figure 3. The required correspondence from the dense matching will be incident to both lines. Search for this correspondence is performed. Finally the colour from the found correspondence is transferred to the new pixel. Due to occlusion there may be several correspondences

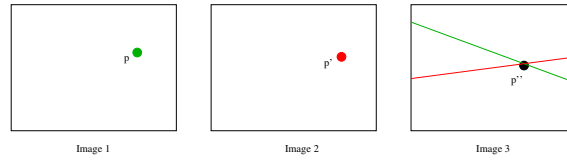


Figure 2: Epipolar Transfer: forward mapping involves intersecting epipolar lines to locate a point in the new view

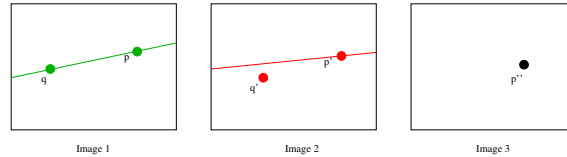


Figure 3: Epipolar Transfer: backward mapping requires search for a consistent correspondence along epipolar lines

incident to both lines and to decide between them the epipole locations are used to determine depth ordering and therefore visibility [6]. Thus, forward mapping is replaced by backward mapping but at the expense of search.

Item 2 above is also dealt with in part by [2] where the view is specified in a more direct way, but still requires the manual selection of correspondences. The location of the desired virtual viewpoint is manually specified by its projection in the original images, and the image plane defined by selecting four correspondences. Our alternative is described in Section 4 which allows direct view specification for epipolar transfer.

The degeneracies apparent in epipolar transfer are well documented [9]. By considering Figure 2 we see that if the epipolar lines are coincident then the intersection is not defined and poorly localised for near coincidence. This case occurs when the *trifocal plane* (the plane defined by the three camera centres) intersects the new image plane. Because an image has a finite extent, this degeneracy is only visible for some virtual views. A completely degenerate case is when the three camera centres are collinear: the trifocal plane is not uniquely defined and therefore intersects the new image everywhere.

### 3.2 Line of Sight

Recent work by Irani et al. [7] develops a ray-tracing like approach to new view synthesis. They consider: what does the scene look like from a scene point? This question can address directly the backward mapping problem since it asks: what colour should this pixel be? The approach is framed in terms of plane-plus-parallax [8] (a representation of visual geometry), with view synthesis from many ( $\approx 10+$ ) plane registered images. We briefly describe the constructions and relate it to epipolar transfer.

A pixel in the new view and the new camera centre define a ray in space. This ray, or *line of sight* may intersect the scene in several points. The idea is that, knowing the image of the new camera centre in each real view (by manual specification) and the plane registering homographies, the images of a line of sight can be used to predict the appearance of each pixel in the new view. Figure 4 illustrates the construction.

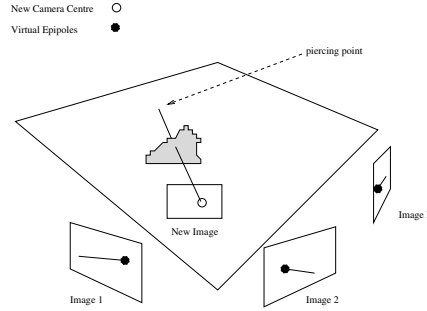


Figure 4: The ray through the pixel of the new image is imaged as a line of sight in the original images.

They observe that a visible scene point is characterised by a consistent colour at a corresponding position along the imaged lines of sight. Due to parallax relative to the plane the colour at any particular position along an imaged line of sight is view dependent except at the corresponding visible points. Thus their work also bears strong similarities to the *space-carving* work of [11].

Their work makes two novel contributions: (i) a direct backward mapping algorithm, (ii) a method of NVS without predetermined correspondence. In order to make the search for colour consistency reliable many images are required (usually ten or more), with results degrading as the number decreases. We see that traditional correlation based correspondence across few images is replaced with pixel comparisons across many.

The line of sight method requires the specification of the images of the new camera centre (virtual epipoles:  $e_{31}, e_{32}$ ) and the homographies between new and real images  $H_{13}$   $H_{23}$ . The ‘collection of fundamental matrices’ required for epipolar transfer e.g.  $F_{32}$  and  $F_{31}$  can be obtained through:

$$F_{31} = [e_{31}]_{\times} H_{13} \quad F_{32} = [e_{32}]_{\times} H_{23}. \quad (2)$$

This method is therefore equivalent, in terms of transfer, to Laveau and Faugeras’ backward mapping [2].

### 3.3 Trifocal Transfer

NVS by trifocal transfer was introduced by Avidan and Shashua [4]. Given two real images (1 and 2) and the tensor for views 1, 2, 3 then transfer is accomplished by transferring (forward mapping) all corresponding pixels to the new view, view 3. One way of obtaining the required tensor is by manually identifying a number of correspondences across all three views. Section 4 describes the alternative offered by the *seed tensor* [4] and in Section 3.4 we outline a backward mapping method for trifocal transfer.

Forward mapping transfer is accomplished using the tensor constraint equation to solve for the point location  $p''$  given a correspondence  $p \leftrightarrow p'$ . We give one possible method of transfer from [9] as follows:

$$p''^k = p^i l'_j T_i^{jk} \quad p'^j l'_j = 0 \quad (3)$$

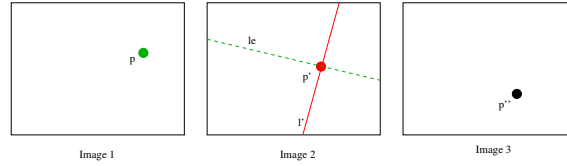


Figure 5: Trifocal transfer by forward mapping: each correspondence along with the trifocal tensor defines a point in the new view

i.e. for each point in the left image  $p$  and a line  $l'$  through its corresponding point  $p'$  use the transfer equation to obtain  $p''$ . The line in the second image may be any line through the point although we choose the line perpendicular to the epipolar line in order that the transfer is more stable. Trifocal transfer does not suffer the singularities evident in epipolar transfer.

### 3.4 Trifocal Transfer by Backward Mapping

Trifocal transfer has proven to be a useful device for NVS. Current implementations produce results that are marred by the simple but problematic forward mapping method. We now describe our backward mapping method for trifocal transfer. The transfer process is inverted to define search lines similar to that described in Section 3.1. We show how this is done given the trifocal tensor.

We denote the trifocal tensor between views  $(3, 1, 2)$  to be  $T_k^{ij}$  and a triplet correspondence as  $(p \leftrightarrow p' \leftrightarrow p'')$ . For a pixel in the third view  $p''$  we obtain a matrix  $J^{ij}$  which defines a correlation between views 1 and 2:

$$J^{ij} = T_k^{ij} p''^k \quad (4)$$

We can think of this correlation being induced by the ray through the third camera centre and the point  $p''$ . The left and right null spaces  $l_0, l'_0$  of  $J^{ij}$  correspond to the images of this ray in the first and second views respectively (compare Section 3.2). Figure 6 illustrates the geometry.

Thus, each point on  $l_0$  may be transferred to a point on  $l'_0$  according to Equation 5. This will only agree with the underlying correspondences for particular point pairs (those that are images of a point on the ray in Figure 6),

$$p^i l_{0i} = 0 \quad p'^j l'_{0j} = 0 \quad p'^j = J^{ij} s_i \quad (5)$$

where,  $s_i$  is any line (except for  $l_{0i}$  – the nullspace) through the point  $p^i$ . This gives the necessary incidence and correspondence relations for performing backward mapping:

- 1: **for** each pixel in novel view **do**
- 2:   compute  $J^{ij}$  and obtain epipolar lines as left and right nullspaces
- 3:   **for** each correspondence on epipolar lines (in visibility order) **do**
- 4:     **if** compatible with equation 5 **then**
- 5:       transfer colour to pixel in novel view
- 6:     **end if**
- 7:   **end for**
- 8: **end for**

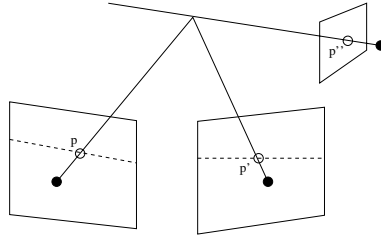


Figure 6: Trifocal transfer by backward mapping: search is performed for a correspondence incident to the ray as shown.

## 4 View Specification

One aspect of NVS that has received less attention than transfer is how to specify the novel viewpoint. Most implementations described in the literature do not pursue this issue and instead rely on indirect specification. This is usually via user selection of a number of correspondences across views. This is obviously not ideal from a user’s point of view; it is difficult to know what view to expect, it is tedious and often an unrealistic warped image is obtained because all degrees of freedom must be selected.

This section describes direct methods for view specification. As one of the novel contributions of this paper we derive the necessary epipolar constraints required to specify a Euclidean camera motion in the scene relative to one of the original views. Avidan and Shashua [4] provide a similarly attractive system of virtual camera specification for trifocal transfer in terms of a Euclidean camera motion and a *seed tensor*.

For Euclidean view specifications it is necessary either to calculate the calibration matrices or, because we are not interested in ground truth but plausibility, assume reasonable calibration matrices. Here we assume a calibration matrix  $K$  common to both views in which the principal point is located at the image centre and the focal length is equal to the image width.

### 4.1 Trifocal View Specification

The trifocal tensor is generally defined for a set of three distinct views, but the *seed tensor* of two views may be constructed. This tensor is simply an extension of the fundamental matrix and is the tensor for views  $(1, 2, 2)$ . Given a pose for the novel view  $(R, t)$  the relevant tensor is computed by updating the seed tensor. The update for the canonical cameras is:

$$\beta_i^{jk} = R_i^k \alpha_i^{jl} - t^k R_i'^j \quad (6)$$

where  $\alpha_i^{jl}$  is the seed tensor (generalised fundamental matrix),  $\beta_i^{jk}$  is the tensor valid for views  $(1, 2, 3)$  where view 3 is rotated and translated relative to view 2. The rotation matrix,  $R'$ , between the two real views is required and may be obtained by standard decomposition of the fundamental matrix.

Therefore, given a dense correspondence between two real views and the fundamental matrix the tensor for a novel camera pose may be computed.



## 4.2 Epipolar View Specification

We consider how the fundamental matrices are obtained for two real views and a Euclidean specified virtual camera pose. The fundamental matrix between views 1 and 2,  $F_{21}$  is upgraded to the status of the essential matrix by  $E = K^T F_{21} K$ . We move our virtual camera according to  $(R, t)$  relative to the second view such that  $E_{21} = [t']_{\times} R'$  and  $E_{32} = [t]_{\times} R$ , and further we obtain the following epipolar constraint:

$$\begin{aligned}
 E_{31} &= [t + Rt']_{\times} RR' \\
 \Rightarrow E_{31} &= [t]_{\times} RR' + [Rt']_{\times} RR' \\
 \Rightarrow E_{31} &= E_{32}R' + R[t']_{\times} R' \\
 \Rightarrow E_{31} &= E_{32}R' + RE_{21}
 \end{aligned} \tag{7}$$

Thus, given the fundamental matrix between the original views we can obtain the fundamental matrices required for epipolar transfer given the desired motion  $(R, t)$ . The trifocal tensor may also be obtained by combining fundamental matrices<sup>1</sup> thus making this view specification method applicable to either transfer method.

## 5 Implementation and Results

We have concentrated on producing a new view from two real images only, in order to facilitate a comparison. Epipolar and trifocal transfer have been implemented using both forward and backward mapping. All novel views are specified according to Section 4. Figure 7 shows two novel views from the generated movies of a lab scene for each combination of transfer and mapping method. We have chosen a scene containing multiple objects with varying depth and therefore scope for parallax and occlusion.

Row (a) of 7 shows the original image pair, rows (b) and (c) show forward mapping for epipolar and trifocal transfer respectively, rows (d) and (e) show backward mapping for epipolar and trifocal transfer respectively. Backward mapping is clearly favourable in terms of rendering quality. The lesser trifocal plane degeneracy can be clearly seen as a diagonal streak in the resulting novel views using epipolar transfer regardless of the mapping method.

The dense matching used is not completely dense because of difficulties in matching low textured regions. This is the reason for more ‘holes’ than usual in the resulting images. We distinguish between ‘holes’ due to finite resolution, which backward mapping corrects, and holes due to occluded regions, which are not handled. Some incorrect matching also leads to holes in regions that are not occluded in both mapping methods.

The indoor scene images were taken with a handheld digital camera and have a resolution of  $640 \times 480$ . The preprocessing consists of dense matching and robust computation of the fundamental matrix. On a 1GHz processor the forward mapping algorithms produced novel views at a rate approximately 1Hz. Backward mapping takes considerably longer (several minutes) and was implemented using disparity constraints from forward mapping to speed up the process.

---

<sup>1</sup>The trifocal tensor may be obtained from the fundamental matrices only if the camera centres are not collinear.

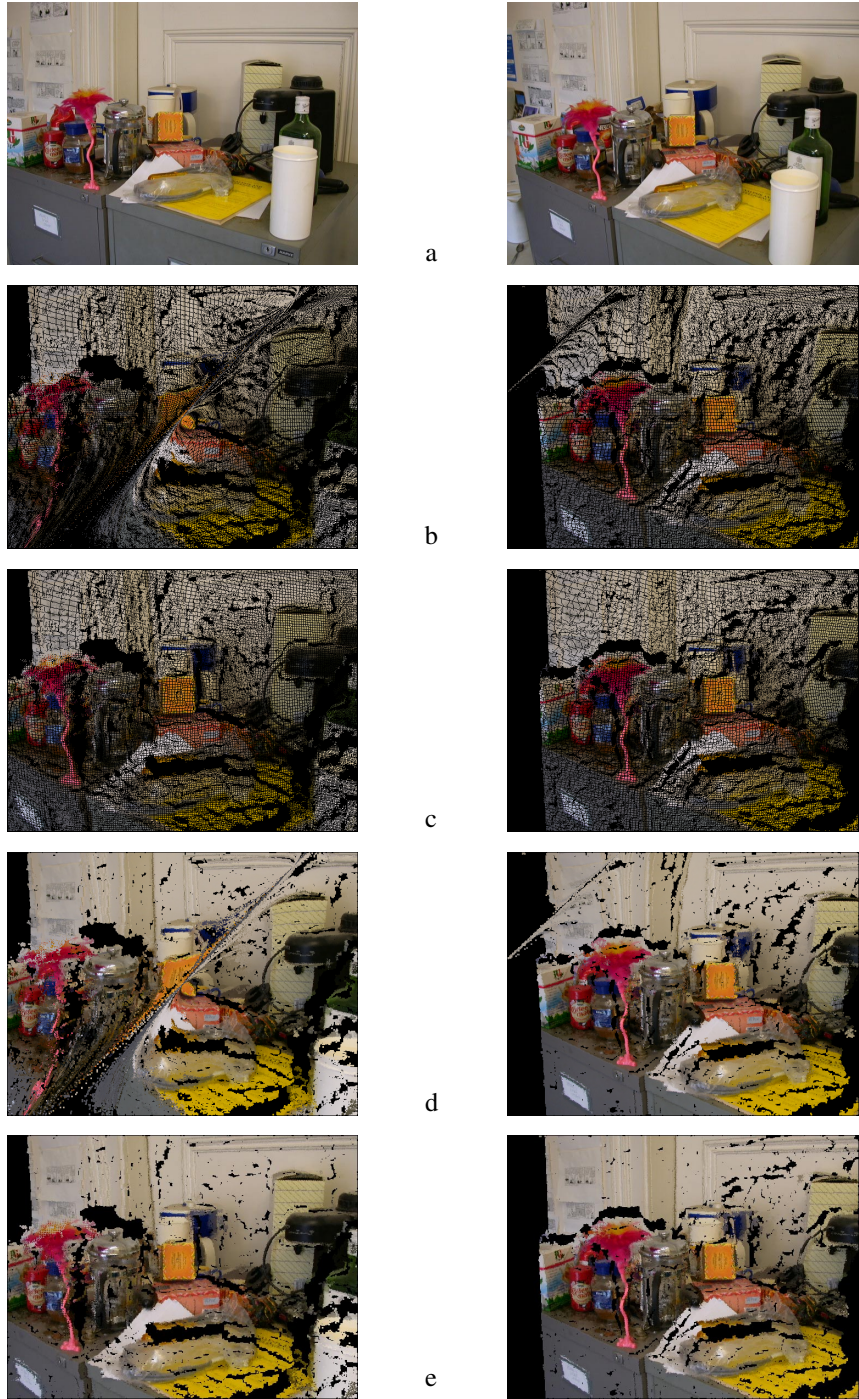


Figure 7: (a) original image pair, Generated novel views by transfer method: (b) forward epipolar, (c) forward trifocal, (d) backward epipolar, (e) backward trifocal.



## 6 Conclusion

This paper has placed several transfer methods into a common framework, allowing treatment of view specification and mapping scheme for Novel View Synthesis. We have shown the clear benefits of backward mapping and the utility of direct view specification. In particular, we have derived and implemented a ‘user friendly’ (Euclidean) view specification scheme for epipolar and trifocal transfer where only two weakly calibrated image pairs are assumed. The machinery of backward mapping for trifocal transfer has also been developed, improving rendering quality.

Although not discussed here, backward mapping may offer a more direct way of dealing with partially occluded regions. Our current work will develop this as well as the extension to multiple real views capturing more extensive scenes.

## 7 Acknowledgements

This work is supported by the European Framework 5 grant *EVENTS* [IST-1999-21125]. We would like to thank Naoji Shiroma for his input and discussion.

## References

- [1] O.Faugeras and L.Robert: *What can two images tell us about a third one?*, In Proc. of 3rd European Conference on Computer Vision, Stockholm, Sweden, pp. 485–492, 1994
- [2] S.Laveau and O.Faugeras: *3-D Scene Representation as a Collection of Images and Fundamental Matrices*, Tech. Rep. 2205, Inria, Sophia-Antipolis, France, Feb 1994
- [3] S.M.Seitz and C.R.Dyer: *Physically-Valid View Synthesis by Image Interpolation*, IEEE Workshop on Representation of Visual Scenes, 1995
- [4] S.Avidan and A.Shashua: *Novel View Synthesis in Tensor Space* In Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1034–1040, 1997
- [5] M.Lhuillier: *Towards Automatic Interpolation for Real and Distant Image Pairs*, Technical Report 3619, Inria, February 1999
- [6] Leonard McMillan: *Computing Visibility Without Depth*, Technical Report, University of North Carolina, October 1995, UNC Computer Science Technical Report TR95-047
- [7] M.Irani, T.Hassner, P.Anandan: *What does the scene look like from a scene point?*, European Conference on Computer Vision, Copenhagen, 2002
- [8] A.Criminisi, I.Reid, A.Zisserman: *Duality, Rigidity, Stratified Metrology and Planar Parallax*, European Conference on Computer Vision, Freiburg, June 1998
- [9] Richard Hartley and Andrew Zisserman: *Multiple View Geometry: in computer vision*, Cambridge University Press
- [10] S.Chen and L.Williams: *View Interpolation for image synthesis*, In Proc. SIGGRAPH 1993, pp. 279–288
- [11] K.N.Kutulakos and S.M.Seitz: *A Theory of Shape by Space Carving*, Technical Report TR692, 1998