Reconstructing Reflective and Transparent Surfaces from Epipolar Plane Images

Sven Wanner and Bastian Goldluecke

Heidelberg Collaboratory for Image Processing

Abstract. While multi-view stereo reconstruction of Lambertian surfaces is nowadays highly robust, reconstruction methods based on correspondence search usually fail in the presence of ambiguous information, like in the case of partially reflecting and transparent surfaces. On the epipolar plane images of a 4D light field, however, surfaces like these give rise to overlaid patterns of oriented lines. We show that these can be identified and analyzed quickly and accurately with higher order structure tensors. The resulting method can reconstruct with high precision both the geometry of the surface as well as the geometry of the reflected or transmitted object. Accuracy and feasibility are shown on both ray-traced synthetic scenes and real-world data recorded by our gantry.

1 Introduction

Multi-view stereo methods have made tremendous progress in the past years and offer ever higher accuracy on standard benchmarks like the Middlebury data sets [1]. However, at its core, multi-view stereo and structure from motion rely on the detection of corresponding regions between images, based on the assumption that a scene point looks the same in all views where it is observed - the scene surfaces need to be diffuse reflectors, i.e. Lambertian. While this assumption is completely unrealistic, as a quick look into any natural scene will immediately make obvious, using priors and optimization one can usually obtain robust results at least for surfaces which exhibit only small amounts of specular reflections.

In the presence of partially reflecting surfaces, however, the Lambertian assumption and thus correspondence matching based on comparison of image color completely breaks down. The overlay of information from surface and reflection causes ambiguous information, which leads to a failure of matching based methods, see Figure 1. Furthermore, it seems unlikely that a traditional stereo pair offers enough information to reliably reconstruct the scene structure in such a setting, although there has been some success in restricted scenarios [2].

We therefore propose to analyze reflecting surfaces in a 4D light field, which can be understood as a collection of views with a special structure. What sets a light field apart from standard multi-view imagery is that the views are densely sampled, and thus we can assume the light field to be defined on a continuous space of rays. In particular, it is possible to compute derivatives of tensor fields on this ray space in view point direction.

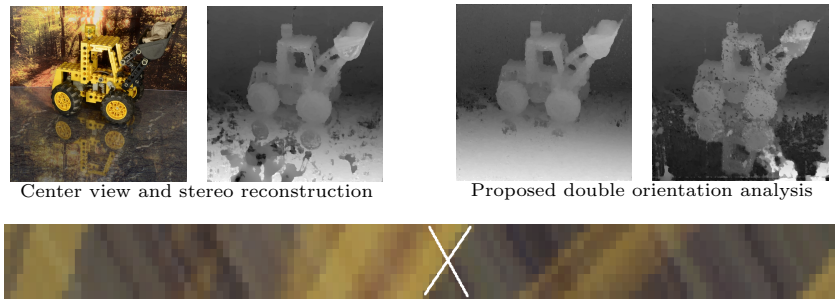


Fig. 1. Reconstructing a mirror. *An algorithm based on the Lambertian assumption cannot distinguish the two signals from mirror plane and reflection and reconstructs erroneous depth for the mirror plane. In contrast, the proposed light field analysis framework correctly separates the data for the mirror plane from the reflection.*

In the present work, we contribute a method to leverage this differential structure to simply and reliably reconstruct the geometry of partially reflective or transparent surfaces, which is a quite hard challenge to achieve from unstructured multi-view data [3–5]. To this end, we employ an image formation model for which reflections or transparencies manifest as overlaid line structures in epipolar plane image space, whose orientation is related to disparity. We then show how to separate these structures into the two different layers (for e.g. mirror and reflection component) using straight-forward pattern analysis methods, see Figure 1. In particular, we use higher order structure tensors for multiple oriented patterns as developed in [6] in order to analyze the directional structure. That way, we derive local estimates for the disparity of both layers, which can be further processed in global optimization schemes. In experiments, we demonstrate the feasibility and accuracy of our method on both ray-traced synthetic as well as real world data sets from a gantry.

2 Related work

While there has been progress in the field of non-Lambertian reconstruction under controlled lighting conditions [7–10], it remains quite hard to generalize the standard matching models to more general reflectance functions if only a set of images under unknown illumination is available. Previous attempts employ a rank constraint on the radiance tensor [3] to derive a discrepancy measure for non-Lambertian scenes. While this improves upon the standard Lambertian matching models and allows to reconstruct surface reflection parameters, the results still somewhat lack in robustness. An interesting alternative approach is Helmholtz stereopsis [11], which makes use of the symmetry of reflectance or Helmholtz reciprocity principle in order to eliminate the view dependency of specular reflections in restricted imaging setups. By alternating light source and camera at two different locations, one can obtain a stereo pair where specularities are exactly identical and thus classical matching techniques can be employed for

(a) Light field structure (b) Light field parametrization

Fig. 2. *(a) A 4D light field is essentially a collection of images of a scene, where the focal points of the cameras lie in a 2D plane. Additional structure becomes visible when one stacks all images along a line of view points on top of each other and considers a cut through this stack (green border above). The 2D image in the plane of the cut is called an* epipolar plane image (EPI)*. (b) Each camera location* (s, t) *in the view point plane* Π *yields a different pinhole view of the scene. The two thick dashed black lines are orthogonal to both planes, and their intersection with the plane* Ω *marks the origins of the* (x, y)*-coordinate systems for the views* (s_1, t) *and* (s_2, t)*, respectively.*

non-Lambertian scenes. Other works try to remove reflection data from images using prior assumptions or user input . However, none of the approaches above allow to infer anything about the geometry of the reflected objects.

The method in [2] conveys similar ideas than ours, but it does not make use of structure tensors and tries to recover the information from only a stereo pair. Consequently, the problem is much more ill-posed and only very restricted settings are analyzed.

Two other works which are closely related to ours are [4, 5]. They both separate a scene which can partially consist of two different layers by considering an epipolar volume constructed from camera motion. At their heart, these works still rely on classical correspondence matching, since they optimize for two overlaid matching models in a nested plane sweep algorithm using graph cuts or semi-global matching, respectively.

In contrast, in our proposed method we do not try to optimize for correspondence. Instead, we build upon early ideas in camera motion analysis [14] and investigate directional patterns in epipolar space. The two layers manifest as overlaid structures, which we investigate with higher order structure tensors [6] as a consequent generalization of [15]. As a result, we obtain a direct continuous method which requires no discretization into depth labels, and which is highly parallelizable and quite fast: a center view disparity map for both layers can be obtained in less than two seconds for a reasonably sized light field, around a hundred times faster than the shortest run-times reported in [5].

3 The structure of epipolar plane images

Light fields and epipolar plane images. A 4D light field or Lumigraph can be imagined as a collection of pinhole views with the same image plane Ω and

focal points lying in a second parallel plane Π. see Figure 2(a). We choose the parametrization detailed in [15]. Coordinates $(s,t) \in \Pi$ define view point locations, and for each such pair, a local (x,y) coordinate system gives the pinhole projection through (s,t) with image plane in Ω, such that a ray $R[s,t,0,0]$ passes through the focal point (s,t) and the center of projection in the image plane, as detailed in Figure 2(b). This parametrization for ray space is slightly different from the standard one for a Lumigraph [16], and inspired by [14].

A light field L is a map which assigns an intensity value (grayscale or color) to each ray. Of particular interest are the images which emerge when the space of rays is restricted to a 2D plane. If we fix for example the two coordinates (y^*,t^*), the restriction L_{y^*,t^*} is the map

$$L_{y^*,t^*} : (x,s) \mapsto L(x,y^*,s,t^*), \tag{1}$$

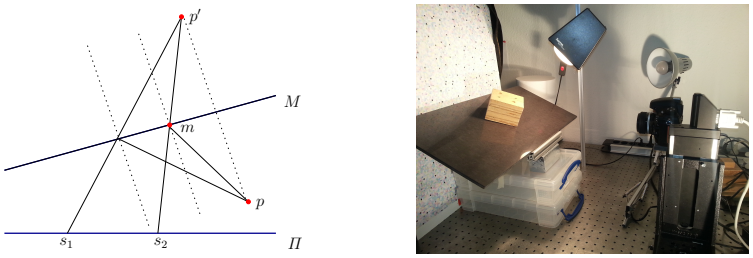
other restrictions are defined in a similar way. Note that L_{s^*,t^*} is the image of the pinhole view with center of projection (s^*,t^*). The images L_{y^*,t^*} and L_{x^*,s^*} are called *epipolar plane images*. They can be interpreted as horizontal or vertical cuts through a horizontal or vertical stack of the views in the light field, see Figure 2(a), and have a rich structure which resembles patterns of overlaid straight lines. Their slope yields information about the scene structure.

EPI structure for opaque Lambertian surfaces. Let $P \in \mathbb{R}^3$ be a scene point. It is easy to show, see e.g. [15], that the projection of P on each epipolar plane image is a straight line with slope $\frac{f}{Z}$, where Z is the *depth of* P, i.e. distance of P to the plane Π, and f the focal length, i.e. distance between the planes Π and Ω. The quantity $\frac{f}{Z}$ is called the *disparity* of P. In particular, the above means that if P is a point on an opaque Lambertian surface, then for all points on the epipolar plane image where the point P is visible, the light field L must have the same constant intensity. This is the reason for the single pattern of solid lines which we can observe in the EPIs of a Lambertian scene.

In [15], this well-known observation was the foundation for a novel approach to depth estimation, which leveraged the structure tensors of the epipolar plane images in order to estimate the local orientation and thus the disparity of the observed point visible in the corresponding ray. While in conjunction with visibility constraints this leads to a certain robustness against specular reflexes, the image formation model implicitly underlying this method is still the Lambertian one, thus the method cannot deal correctly with reflecting surfaces.

EPI structure in the presence of planar reflectors. We now introduce an idealized appearance model for the epipolar plane images in the presence of a planar mirror - a translucent surface is an obvious specialization where a real object takes the place of the virtual one behind the mirror. It is kept simple in order to arrive at a computationally easily tractable model, but we will see that it captures the characteristics of reflective and translucent surfaces reasonably well to be able to cope with real-world data. A similar appearance model was successfully employed in [5].

Let $M \subset \mathbb{R}^3$ be the surface of a planar mirror. We fix coordinates (y^*,t^*) and consider the corresponding epipolar plane image L_{y^*,t^*}. The idea of the



(a) Mirror plane geometry (b) Recording setup

Fig. 3. *(a) Geometry of reflection on a planar mirror. All cameras view the reflections of a scene point p at a planar mirror M as the image of a virtual point p′ which lies* behind *the mirror plane. (b) Our gantry construction we employed to record the real-world light fields.*

appearance model is to define the observed color for a ray at location (x, s) which intersects the mirror at $m \in M$. Our simplified assumption is that the observed color is a linear combination of two contributions. The first is the base color $c(m)$ of the mirror, which describes the appearance of the mirror without the presence of any reflection. The second is the color $c(p)$ of the reflection, where p is the first scene point where the reflected ray intersects the scene geometry, see Figure 3(a). We do not consider higher order reflections, and assume the surface at p to be Lambertian. We also assume the reflectivity $\alpha > 0$ is a constant independent of viewing direction and location. The epipolar plane image itself will then be a linear combination

$$L_{y^*,t^*} = L^M_{y^*,t^*} + \alpha L^V_{y^*,t^*} \tag{2}$$

of a pattern $L^M_{y^*,t^*}$ from the mirror surface itself as well as a pattern $L^V_{y^*,t^*}$ from the virtual scene behind the mirror. In each point (x, s) as above, both constituent patterns have a dominant direction corresponding to the disparities of m and p. The next section shows how to extract these two dominant directions.

4 Double-orientation detection on an EPI

Structure tensors for multi-orientation models. We briefly summarize the theory for the analysis of superimposed patterns described in [6]. A region $R \subset \Omega$ of an image $f : \Omega \to \mathbb{R}$ has orientation $\boldsymbol{v} \in \mathbb{R}^2$ if and only if $f(x) = f(x + \alpha\boldsymbol{v})$ for all $x, x + \alpha\boldsymbol{v} \in R$. Analysis shows that the orientation $\boldsymbol{v}$ is given by the Eigenvector corresponding to the smaller Eigenvalue of the structure tensor [17] of f. However, the model fails if the image f is a superposition of two oriented images, $f = f_1 + f_2$, where f_1 has orientation $\boldsymbol{u}$ and f_2 has orientation $\boldsymbol{v}$. In this case, the two orientations $\boldsymbol{u}, \boldsymbol{v}$ need to satisfy the conditions

$$\boldsymbol{u}^T \nabla f_1 = 0 \text{ and } \boldsymbol{v}^T \nabla f_2 = 0 \tag{3}$$

individually on R.

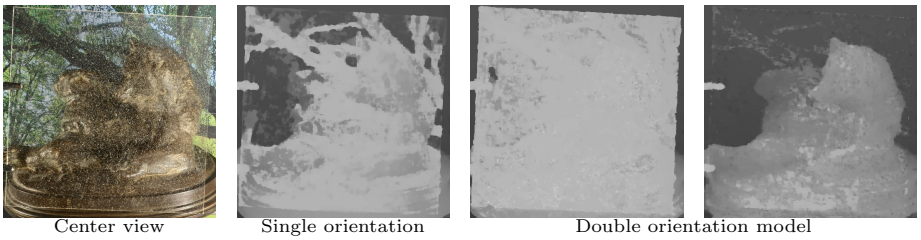


Fig. 4. Reconstructing a transparent surface. *The single orientation model cannot distinguish the two signals from the dirty glass surface and the objects behind it. In contrast, multi-orientation analysis correctly separates both layers.*

Analogous to the single orientation case, the two orientations in a region R can be found by performing an Eigensystem analysis of the second order structure tensor, see [6],

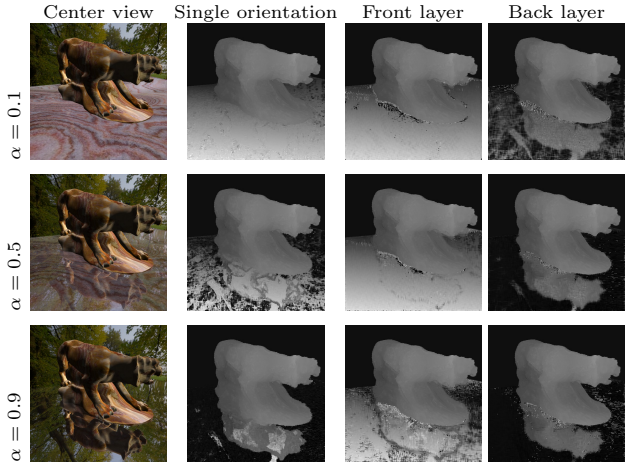
$$\mathcal{T} = \int_R \sigma \begin{bmatrix} f_{xx}^2 & f_{xx}f_{xy} & f_{xx}f_{yy} \\ f_{xx}f_{xy} & f_{xy}^2 & f_{xy}f_{yy} \\ f_{xx}f_{yy} & f_{xy}f_{yy} & f_{yy}^2 \end{bmatrix} \mathrm{d}(x, y), \tag{4}$$

where σ is a (usually Gaussian) weighting kernel on R which essentially determines the size of the sampling window. Since $\mathcal{T}$ is symmetric, we can compute Eigenvalues and Eigenvectors in a straight-forward manner using the explicit formulas in [18]. Analogous to the Eigenvalue decomposition of the 2D structure tensor, the Eigenvector $\boldsymbol{a} \in \mathbb{R}^3$ corresponding to the smallest Eigenvalue of $\mathcal{T}$, the so-called *MOP vector*, encodes the orientations. Indeed, the two disparities are equal to the Eigenvalues λ_+, λ_- of the 2×2 matrix

$$\begin{bmatrix} a_2/a_1 & -a_3/a_1 \\ 1 & 0 \end{bmatrix}, \tag{5}$$

from which one can compute the orientations $\boldsymbol{u} = [\lambda_+ \; 1]^T$ and $\boldsymbol{v} = [\lambda_- \; 1]^T$. The additional material contains source code fragments for the above key steps for easy reimplementation of our method.

Merging the results into a single disparity map. From the steps sketched above, we obtain three different disparity estimates for both the horizontal as well as vertical epipolar images: one from the single orientation model, and two from the double orientation model. It is clear that the closer estimate in the double orientation model will always correspond to the primary surface, regardless of whether it is a mirror or translucent object. Unfortunately, we do not know yet of a reliable mathematical measure which tells us whether the two-layer model is valid or not. We therefore impose a simple heuristic: if at a given point, the disparity values of horizontal and vertical EPIs agree up to a small error for both the primary and secondary orientation, we flag the double orientation model as valid, and choose its contribution in the disparity maps. Otherwise, we choose the estimate from the single orientation model.



Reflection coefficient	Point-wise result				Result after TV-L^2 denoising			
	Single orientation		Double orientation		Single orientation		Double orientation	
	front	back	front	back	front	back	front	back
$\alpha = 0.1$	0.0034	0.7409	0.0078	0.1191	0.0025	0.7392	0.0036	0.09924
$\alpha = 0.3$	0.0236	0.5994	0.0061	0.0349	0.0086	0.6273	0.0032	0.02371
$\alpha = 0.5$	0.0869	0.3735	0.0066	0.0236	0.0252	0.5111	0.0036	0.01377
$\alpha = 0.7$	0.1807	0.1547	0.0101	0.0239	0.1434	0.1821	0.0060	0.01053
$\alpha = 0.9$	0.2579	0.0365	0.0389	0.0473	0.2557	0.0312	0.0249	0.00980

Fig. 5. Influence of reflectivity on accuracy. *The table shows mean squared disparity error in pixels of the single and double orientation model for both the mirror plane as well as the reflection. While the single orientation model shifts from reconstruction of mirror to reflection with growing reflectivity α, the double orientation model can still reconstruct both when even a human observer has difficulties separating them. The images show the point-wise results.*

5 Results

We compare our method primarily to the single orientation method [15] based on the first order structure tensor, which is similar in spirit and an initial step in our algorithm in any case. However, it is clear that any multi-view stereo method will have similar problems than the single orientation method if the underlying model is also the Lambertian world. More results and full-resolution images which would exceed the allowed space of the paper can be found in the additional material. The light fields as well as the complete source code for the method to reproduce the results will be available online.

Synthetic data sets. Figure 5 shows reconstruction accuracy on a synthetic light field with varying amounts of reflectivity α. The scene was ray-traced in a way which exactly fits the image formation model. As expected, the disparity reconstructed with the single orientation model is close to the disparity of the mirror surface if α is small, and close to the disparity of the reflection if α is large. In between, the result is a mixture between the two, depending on whose texture

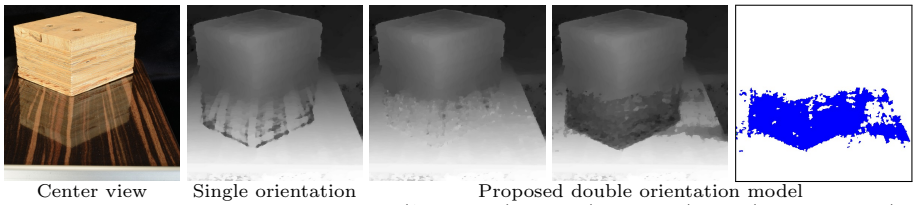


Fig. 6. *In the absence of a structured background, the reflecting surface can of course only reliably be detected where a reflection of a foreground object is visible. The blue region indicates where the double orientation model returns valid results.*

is stronger. In contrast, the double orientation model can reliably reconstruct both reflection as well as mirror surface for the full range of reflectivities α, even when it is already difficult for a human to still observe both.

While the point-wise results are already very accurate, they are still quite noisy and can be greatly improved by adding a small amount of TV-L^2 denoising [19]. We deliberately do not employ more sophisticated global optimization in this step to showcase only the raw output from the model and what is possible at interactive performance levels. For all of the light fields shown, at image resolutions upwards of 512×512 with 9×9 views, the point-wise disparity computation for the whole center view takes less than 1.5 seconds on an nVidia GTX 680 GPU.

Real-world data sets. In Figures 4, 6, and 7, we show reconstruction results for light fields recorded with our gantry, see Figure 3(b). Each one has 9×9 views at resolutions between 0.5 and 1 megapixels. For both reflective and transparent surfaces, a reconstruction of a single disparity based on the Lambertian assumption produces major artifacts and is unusable in the region of the surface. In contrast, the proposed method always produces a very reliable estimate for the primary surface, as well as a reasonably accurate one for the reflected or transmitted objects, respectively. For the results in the figures, we employed a global optimization scheme [20, 15] to reach maximum possible quality, which takes about 3 minutes per disparity map. The same scheme and parameters were used for both methods and all data sets. To show what is possible in near real-time, we also provide the raw point-wise results in the additional material.

The results show that certain appearant limitations of the model are not practically relevant. In particular, different to our simplified model, the reflectivity α is certainly not constant everywhere due to influences of e.g. the Fresnel term, but since all estimates are strictly local and the angular range small, the variations do not seem to impact the final result by much. A stronger limitation, however, is the planarity of the reflecting or transparent surface. We predict that it can be considerably weakened, since the main assumption of the existence of an object "behind" the primary surface (which is of course only virtual in case of a mirror) also holds for more general geometries. However, exploring this direction is left for future work.

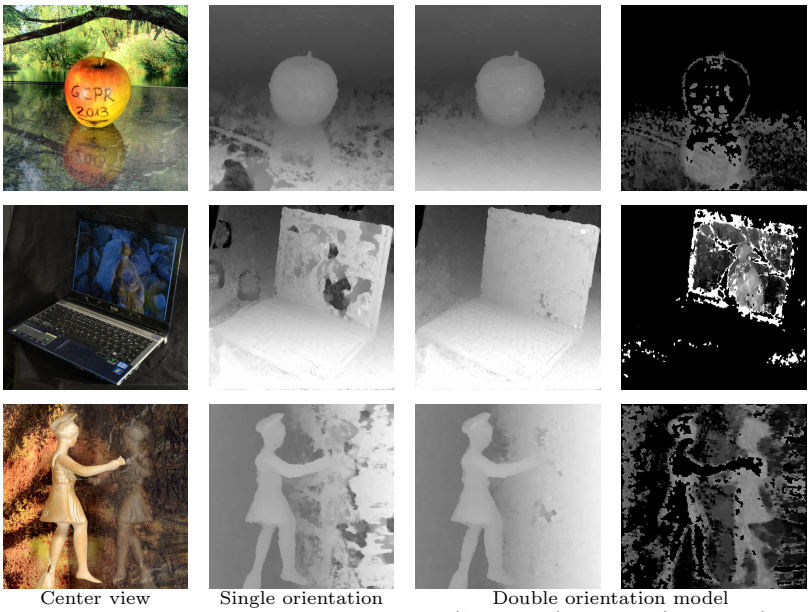


Fig. 7. Reconstructing a mirror. *Like multi-view stereo algorithms, the single orientation model cannot distinguish the two signals from mirror plane and reflection and reconstructs erroneous disparity for the mirror plane. In contrast, the proposed double orientation analysis correctly separates the data for the mirror plane from the reflection. The reflection layer is masked out where the double orientation model does not return valid results as specified in section 4, and the results for this layer have been increased in brightness and contrast for better visibility (raw results and many more data sets can be observed in the additional material).*

6 Conclusion

We have described a method to simultaneously compute disparity maps for both a planar reflective or transparent surface as well as the reflected or transmitted object in a 4D light field structure. In this scenarios, two different oriented structures are overlaid on the epipolar plane images, which we propose to analyze using second order structure tensors. We have demonstrated that this approach leads to reliable disparity estimates which can be computed at interactive speeds. Since the problem of reconstructing reflective or semi-transparent surfaces is hard to solve efficiently with conventional methods based on stereo matching techniques, we believe that the proposed method could turn out to become a foundation for unique applications of 4D light fields and their acquisition technologies.

References

1. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proc. International

Conference on Computer Vision and Pattern Recognition. (2006) 519–528
2. Borga, Knutsson: Estimating multiple depths in semi-transparent stereo images. In: SCIA. (1999)
3. Jin, H., Soatto, S., Yezzi, A.: Multi-view stereo reconstruction of dense shape and complex appearance. International Journal of Computer Vision **63**(3) (2005) 175–189
4. Tsin, Y., Kang, S., Szeliski, R.: Stereo matching with linear superposition of layers. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(2) (2006) 290–301
5. Sinha, S., Kopf, J., Goesele, M., Scharstein, D., Szeliski, R.: Image-based rendering for scenes with reflections. ACM Transactions on Graphics (Proc. SIGGRAPH) **31**(4) (2012) 100:1–100:10
6. Aach, T., Mota, C., Stuke, I., Muehlich, M., Barth, E.: Analysis of superimposed oriented patterns. IEEE Transactions on Image Processing **15**(12) (2006) 3690–3700
7. Davis, J., Yang, R., Wang, L.: BRDF invariant stereo using light transport constancy. In: Proc. International Conference on Computer Vision. (2005)
8. Alldrin, N., Zickler, T., Kriegman, D.: Photometric stereo with non-parametric and spatially-varying reflectance. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2008)
9. Ruiters, R., Klein, R.: Heightfield and spatially varying BRDF reconstruction for materials with interreflections. Computer Graphics Forum (Proc. Eurographics) **28**(2) (2009) 513–522
10. Goldman, D., Curless, B., Hertzmann, A., Seitz, S.: Shape and spatially-varying BRDFs from photometric stereo. IEEE Transactions on Pattern Analysis and Machine Intelligence **32**(6) (2010) 1060–1071
11. Zickler, T., Belhumeur, P., Kriegman, D.: Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. International Journal of Computer Vision **49**(2–3) (2002) 215–227
12. Levin, A., Zomet, A., Weiss, Y.: Separating reflections from a single image using local features. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2004)
13. Levin, A., Weiss, Y.: User assisted separation of reflections from a single image using a sparsity prior. IEEE Transactions on Pattern Analysis and Machine Intelligence (2007)
14. Bolles, R., Baker, H., Marimont, D.: Epipolar-plane image analysis: An approach to determining structure from motion. International Journal of Computer Vision **1**(1) (1987) 7–55
15. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4D light fields. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2012) 41–48
16. Gortler, S., Grzeszczuk, R., Szeliski, R., Cohen, M.: The Lumigraph. In: Proc. SIGGRAPH. (1996) 43–54
17. Bigün, J., Granlund, G.H.: Optimal orientation detection of linear symmetry. In: Proc. International Conference on Computer Vision. (1987) 433–438
18. Smith, O.: Eigenvalues of a symmetric 3x3 matrix. Communications of the ACM **4**(4) (1961) 168
19. Chambolle, A.: An algorithm for total variation minimization and applications. Journal of Mathematical Imaging and Vision **20**(1-2) (2004) 89–97
20. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM Journal on Imaging Sciences (2010)