Shape from Light Field Meets Robust PCA*

Stefan Heber[1] and Thomas Pock[1,2]

[1] Institute for Computer Graphics and Vision
Graz University of Technology
[2] Safety & Security Department
AIT Austrian Institute of Technology

Abstract. In this paper we propose a new type of matching term for multi-view stereo reconstruction. Our model is based on the assumption, that if one warps the images of the various views to a common warping center and considers each warped image as one row in a matrix, then this matrix will have low rank. This also implies, that we assume a certain amount of overlap between the views after the warping has been performed. Such an assumption is obviously met in the case of light field data, which motivated us to demonstrate the proposed model for this type of data. Our final model is a large scale convex optimization problem, where the low rank minimization is relaxed via the nuclear norm. We present qualitative and quantitative experiments, where the proposed model achieves excellent results.

Keywords: light field, nuclear norm, low rank.

1 Introduction

One of the most studied problems in Computer Vision (CV) is stereo. Given two or more images, taken from a static scene, but from different viewpoints, stereo algorithms try to find points in the different images, that correspond to the same scene point. Therefore the problem is also denoted as the correspondence problem. One distinguishes between local (*cf.* [30]) and global methods (*e.g.* [4]). In both cases one has to define a matching term, that measures the similarity between two image positions. In the case of two-frame stereo this matching term measures how well positions in the reference view match certain positions in the warped view. In the general case of multi-view stereo [25,12,31], proposed methods usually only match the different warped views with a predefined reference view. By increasing the number of matchings, *i.e.* also among the warped views, one could increase the robustness to various problems, which disturb the matching. Such problems can depend on the scene itself, like *e.g.* due to depth discontinuities, specularity, reflections, *etc.*, but could also represent problems of the used image capturing device *e.g.* pixel errors, sensor noise, *etc.*. In order

* This research was supported by the FWF-START project *Bilevel optimization for Computer Vision*, No. Y729 and the Vision+ project *Integrating visual information with independent knowledge*, No. 836630.

D. Fleet et al. (Eds.): ECCV 2014, Part VI, LNCS 8694, pp. 751–767, 2014.

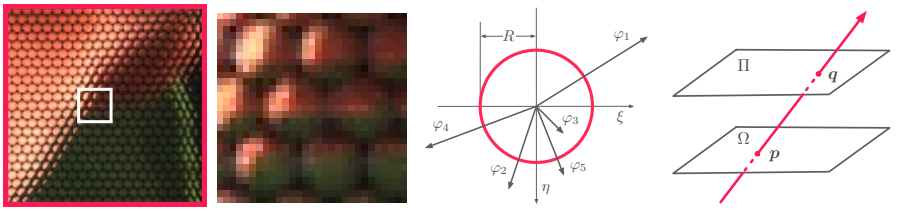

Fig. 1. (a) and (b) show two closeup views of the raw image data captured with a *plenoptic 1.0* camera. One can clearly see the effect of the micro-lens array, where each micro-lens splits incoming light into rays of different directions. Each of those light rays then hits the behind placed sensor at a slightly different location. (c) shows a sketch of the parametrization used in the RPCA light field model (*cf*. (10)). (d) is a visualization of the so-called two plane parametrization of the light field.

to fully exploit the potential of increasing the stability and the accuracy of the reconstruction an algorithm must be able to perform an all vs. all matching. Inspired by low rank models like robust Principal Component Analysis (RPCA) [7], the proposed method tackles this problem by introducing a novel matching term, which globally measures how well the different views can we warped to a common warping center. Hence, this global matching term defines a measure on the complete set of warped images, and we will see that this can also be interpreted as an all vs. all matching between the involved views. To the best of our knowledge such a stereo-model with a global matching term has not been proposed before.

One extreme case of a multi-view system is light field imaging, where a large amount of highly overlapping views are available. One way to capture a light field is by using a so-called plenoptic camera, where the different views are noise and highly aliased. We will show in the experiment section that the proposed method performs especially well for this type of data.

In what follows, we will first give a brief introduction to light field imaging in Section 1.1, followed by a short overview of RPCA [7] in Section 1.2.

1.1 Light Fields

Computer Vision traditionally focuses about extracting information out of images captured with traditional cameras. Nowadays there exist a variation of unconventional cameras, that do not capture traditional images. Among those devices are for instance cameras with coded apertures [20,35], multi-view systems [40], or plenoptic cameras [1,24,23]. All of those cameras have in common, that each point at the sensor sums over a set of light rays, where the optic defines the mapping between light rays and the sensor position. Hence, one can distinguish between different image capturing devices by analyzing the combination of light rays, that hit certain points at the sensor. The so-called light

field [21] is a representation of all light rays, that hit the sensor plane from different directions. Thus, the light field can be seen as a common denominator across different types of cameras. In a pinhole camera for instance each sensor records one light ray. The sensor in a conventional camera records the integral of light rays over the lens aperture. A plenoptic camera seeks to capture the complete light field, *i.e.* it tries to record all light rays hitting certain points at the sensor separately. An image captured with a plenoptic camera can be considered as not static, *i.e.* that the image can be modified after is has been taken, w.r.t. viewpoint, focus and depth of field. Capturing the different light rays per sensor position is achieved by placing a microlens-array in front of the sensor of a traditional camera (*cf.* Figure 1(a) and 1(b)). Note, that a plenoptic camera can not capture the complete light field, it can only capture a light field with a certain directional and spatial resolution. The directional resolution depends on the number of pixels which capture the image of one microlens, and the spatial resolution depends on the overall size of the sensor. The basic concept of a plenoptic camera was first proposed by Lippmann [22] in 1908 and has then been developed and improved [11,13,15], but plenoptic cameras became feasible not until recent years [24,23,28]. The reason is simple due to the fact, that adequate high quality microlens arrays, and high resolution sensors, were not available till recent years.

Mathematically, a light field $\hat{L}$ is a 4D function, which is usually parametrized via the so called two plane parametrization (*cf.* Figure 1(d))

$$\hat{L} : \Omega \times \Pi \to \mathbb{R}, \qquad (\boldsymbol{p}, \boldsymbol{q}) \mapsto \hat{L}(\boldsymbol{p}, \boldsymbol{q}) \tag{1}$$

where $\boldsymbol{p} := (x, y)^{\mathrm{T}}$ denotes a point in the image plane $\Omega \subset \mathbb{R}^2$ and $\boldsymbol{q} := (\xi, \eta)^{\mathrm{T}}$ denotes a points in the lens plane $\Pi \subset \mathbb{R}^2$. There are different ways to visualize the 4D light field. One way is to fix two coordinates and vary over the remaining two. The most useful representations are epipolar and sub-aperture images. An epipolar image is obtained by fixing one spatial coordinate of $\boldsymbol{p}$ and one directional coordinat of $\boldsymbol{q}$. A sub-aperture image is an images where the directional component $\boldsymbol{q}$ is kept constant, and one varies over all spatial positions $\boldsymbol{p}$. Sub-aperture images can also be seen as images extracted out of the light field with slightly different viewpoints, but parallel to a common image plane. Such images clearly show the connection between light fields and multi-view stereo systems, and thus this representation will be used in the proposed model.

Light fields have been used for various image processing applications, such as digital refocusing [18,23], extending the depth of field [23], image super-resolution [3,37], and depth estimation [36,16,32]. In the case of depth estimation, the method proposed by Wanner *et al.* [36] makes use of the epipolar representation of the light field, where they additionally enforce global visibility constraints. Heber *et al.* [16] proposed a method, which uses the sub-aperture representation of the light field, where all views are matched against the center view. Finally, Tao *et al.* [32] suggested a method, that combines the defocus and correspondence depth cues.

1.2 Robust Principal Component Analysis

In many practical situation it is well justified to assume that the given data lies approximately on a low dimensional linear subspace [14,10,2,33]. This means, if data points are stacked as column or row vectors of a matrix M, then M should have low rank. This leads to the following model

$$M = L_0 + E_0\,, \tag{2}$$

where L_0 is assumed to have low rank and E_0 is a perturbation matrix representing the noise. This property has been exploited by classical Principal Component Analysis (PCA) [14,17,19], which solves the following minimization problem

$$\begin{aligned} \underset{L,E}{\text{minimize}} \quad & \|E\|_F \\ \text{subject to} \quad & \operatorname{rank}(L) \leqslant r \\ & M = L + E \end{aligned} \tag{3}$$

where $\|.\|_F$ denotes the Frobenius norm. In problem (3) it is assumed that the entries in E are independent and identically distributed (iid) according to an isotropic Gaussian distribution. In this case PCA provides an optimal estimate to L_0. Also note, that problem (3) can be solved exactly using the singular value decomposition (SVD) of M.

PCA is used extensively for data analysis and dimension reduction, but it may fail in general if the assumptions about the perturbation matrix E are not met, *i.e.* that a few corrupted entries in M, which significantly deviate from the true solution, can lead to an estimate L, that is far away from L_0. Thus, PCA is only effective against small Gaussian noise, but it is highly sensitive to even sparse errors of high magnitude. Such errors are quite common in many applications due to corrupted data, sensor failures, *etc.* Also note, that the rank r of M needs to be known a priori, which is usually not the case in real-world applications.

An algorithm that efficiently extract the principal components of such data even in the presence of large errors was proposed by Candès *et al.* [7]. They assume the rank of L to be unknown, and hence formulate a matrix rank minimization problem, where they want to find the lowest rank that generates M when added with unknown sparse outliers. More specifically, they consider the following combinatorial optimization problem

$$\underset{L,E}{\text{minimize}} \quad \operatorname{rank}(L) + \mu\|E\|_0 \quad \text{subject to} \quad M = L + E\,, \tag{4}$$

where $\|.\|_0$ denotes the number of non zero entries (ℓ^0 norm). Problem (4) is NP hard, and hence can not be solved efficiently. Thus they relax the problem by using the nuclear norm and the ℓ^1 norm to encourage low rankness and sparsity, respectively. Note, that the ℓ^1 norm is the largest convex function below $\|.\|_0$, and the nuclear norm, denoted as $\|.\|_*$, is the largest convex function below the rank function. The nuclear norm of a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ is defined as

$$\|X\|_* = \sum_{i=1}^{n} \sigma_i(X) \quad \text{with} \quad n := \min\{n_1, n_2\}\,, \tag{5}$$

where $\sigma_1(X) \geqslant \sigma_2(X) \geqslant \ldots \geqslant \sigma_n(X) \geqslant 0$ are the singular values of X. By considering the definition of the nuclear norm one sees, that this norm can be interpreted as the ℓ^1 norm of the vector of singular values of X *i.e.* the ℓ^1 norm of the spectrum. This also shows the close relation to compressed sensing.

By using these relaxations, the problem of separating the low rank component from a sparse component can be cast into a convex problem, denoted as Principal Component Pursuit (PCP) problem

$$\underset{L,E}{\text{minimize}} \quad \|L\|_* + \mu\|E\|_1 \quad \text{subject to} \quad M = L + E\,. \tag{6}$$

Also note, that problem (6) can be recast as a semidefinite program (SDP). The method termed Robust Principal Component Analysis (RPCA) performs well in practice and provides the low rank solution, even if up to a third of the observations are grossly corrupted.

Inspired by the RPCA [7], we formulate a holistic matching term for a multi-view stereo model, where we warp images in a way to minimize the rank of the set of warped images. The proposed model assumes that the different images provide a certain amount of overlap, which is particularly true for light field data. Thus, we will demonstrate it on this type of data, but the proposed model is not limited to the light field setting.

It is worth mentioning, that similar ideas have been used by Yigang Peng *et al.* [26] to calculate misalignments of a set of images. However, their method is limited to one global domain transformation, *i.e.* the misalignments between images are modeled as transformations from a finite dimensional group, that has a parametric representation (*e.g.* the similarity group $\mathrm{SE}(2) \times \mathbb{R}_+$, the 2D affine group $\mathrm{Aff}(2)$, or the planar homography group $\mathrm{GL}(3)$).

Contribution

The contribution of this paper is threefold. First, we propose a novel variational multi-view stereo model based on low rank minimization, where the main contribution relies in the theoretical novelty of the RPCA matching term, which can be interpreted as an all vs. all matching term. Second, we present an extension of the proposed model to simultaneous image super-resolution on all low rank components. Third, we show how to apply the model to the light field setting, yielding the RPCA light field model. We then provide a simple optimization scheme, which is describe in detail in Section 3. Final we also present qualitative and quantitative experiments on synthetic and real-world data in Section 4.

2 RPCA Matching

In this section we describe the proposed model, which includes the novel RPCA matching term. The main idea of the model is to globally measure how well a set of warped images is aligned, *i.e.* our model warps images of different viewpoints to a predefined warping center in a way, such that the set of warped images

can be split up into a low rank component and into an sparse component. In mathematical terms the combinatorial problem, which we want to solve can be formulated as follows

$$\begin{aligned} \underset{L,S,u}{\text{minimize}} \quad & \mu \operatorname{rank}(L) + \lambda \|S\|_0 + \mathcal{R}(u) \\ \text{subject to} \quad & I(u) = L + S \end{aligned} \tag{7}$$

where λ, $\mu > 0$ are modeling parameters, and $\mathcal{R}(u)$ denotes a convex regularization term on the disparity variables u. Moreover, $I(u) \in \mathbb{R}^{M \times mn}$ denotes the set of M warped images of size $m \times n$, where each row of $I(u)$ represents one image. The main idea of the proposed model is to estimate a piecewise smooth disparity map, that allows to warp the input images in such a way, that the set of warped images $I(u)$ can be split up into a low rank component L and into a sparse outlier component S. Unfortunately, with the ℓ^0 minimization on S and on the spectrum of L the problem is NP-hard. Note, that the rank of L equals the ℓ^0 norm of the spectrum of L.

Now we follow RPCA [7] and consider a convex relaxation of the above problem, *i.e.* we will relax the sparsity assumption of S with the ℓ^1 norm, and we will model the low rank constraint of L with the nuclear norm. This leads to the following problem

$$\begin{aligned} \underset{u,L,S}{\text{minimize}} \quad & \mu \|L\|_* + \lambda \|S\|_1 + \mathcal{R}(u) \\ \text{subject to} \quad & I(u) = L + S \end{aligned} \tag{8}$$

By eliminating the constraint we then obtain

$$\underset{u,L}{\text{minimize}} \quad \mu \|L\|_* + \lambda \|L - I(u)\|_1 + \mathcal{R}(u)\,. \tag{9}$$

Compared to models with a pointwise or local data fidelity term, this model now globally measures how well the warped images match with each other, *i.e.* all views are considered equivalently important, or in other words this can be seen as an all vs. all matching. Moreover it adjusts the warped views to cope with sparse outliers which are present due to *e.g.* occlusion, specularity, or pixel errors. Also note that we do not define the matching between different views explicitly. The proposed model uses an implicit all vs. all matching via the nuclear norm. Problem (9) can also be interpreted as stereo reconstruction with simultaneously denoising the warped images. So it can be seen as solving jointly the stereo and denoising problem. In order to obtain a convex model we will use first order Taylor approximations to linearize the warped images in a final step.

2.1 Application to Light Field Imaging

In the case of light field data, we will warp so-called sub-aperture images to a predefined warping center, *e.g.* the center view of the light field. Assuming

an ideal *plenoptic 1.0 camera* and using a similar notation as in [16], the sub-aperture images are defined as follows (*cf*. Figure 1(c))

$$\tilde{I}_i(\tilde{u}) := \left(\hat{L}\left(\boldsymbol{p} - \tilde{u}(\boldsymbol{p})\frac{\varphi_i}{R}, \varphi_i\right)\right)_{\boldsymbol{p}\in\hat{\Omega}}, \quad \text{with} \quad 1 \leqslant i \leqslant M, \tag{10}$$

where M denotes the number of different sub-aperture images, φ_i is the directional offset of the i^{th} sub-aperture image, $\hat{\Omega} := \{(x,y)^{\mathrm{T}} \in \mathbb{N}_0^2 \,|\, x < n,\, y < m\}$ is the discrete image grid, and $\tilde{u} : \hat{\Omega} \to \mathbb{R}$ is the disparity between the warping center and images with a predefined directional offset distance R. By reshaping the images $\tilde{I}_i(\tilde{u})$ as row vectors, one can define the matrix $I(u) \in \mathbb{R}^{M\times mn}$, where each row represents one sup-aperture image as defined in (10). For this purpose we define a vectorization operator vec(.), which transforms an image in matrix representation to a column vector in row major representation, *i.e*. that the i^{th} row of $I(u)$ in problem (9) is now equivalent to $\mathrm{vec}(\tilde{I}_i(\tilde{u}))^{\mathrm{T}}$.

In order to obtain a convex model we have to linearize the warped images. Therefore, we use a first order Taylor approximation for each sub-aperture image at the position $\tilde{u}_0$

$$\hat{L}\left(\boldsymbol{p} - \tilde{u}_0(\boldsymbol{p})\frac{\varphi_i}{R}, \varphi_i\right) + (\tilde{u}(\boldsymbol{p}) - \tilde{u}_0(\boldsymbol{p}))\,\frac{\|\varphi_i\|}{R}\,\nabla_{-\frac{\varphi_i}{\|\varphi_i\|}}\hat{L}\left(\boldsymbol{p} - \tilde{u}_0(\boldsymbol{p})\frac{\varphi_i}{R}, \varphi_i\right), \tag{11}$$

where $\nabla_{\boldsymbol{v}}$ denotes the directional derivatives with direction $[\boldsymbol{v}, \boldsymbol{0}]$. To simplify notation we define $\tilde{A}_i$ and $\tilde{B}_i \in \mathbb{R}^{m\times n}$ similar as in [16]

$$\tilde{A}_i := \left(\frac{\|\varphi_i\|}{R}\,\nabla_{-\frac{\varphi_i}{\|\varphi_i\|}} L\left(\boldsymbol{p} - \tilde{u}_0(\boldsymbol{p})\frac{\varphi_i}{R}, \varphi_i\right)\right)_{\boldsymbol{p}\in\hat{\Omega}}, \tag{12}$$

$$\tilde{B}_i := \left(L\left(\boldsymbol{p} - \tilde{u}_0(\boldsymbol{p})\frac{\varphi_i}{R}, \varphi_i\right)\right)_{\boldsymbol{p}\in\hat{\Omega}}. \tag{13}$$

Now we set $b_i = \mathrm{vec}(\tilde{B}_i)$ and $A_i = \mathrm{diag}(\mathrm{vec}(\tilde{A}_i))$, which allows to rewrite problem (9) as the following convex optimization problem

$$\underset{u,L}{\text{minimize}} \quad \mu\|L\|_* + \lambda\sum_{i=1}^{M}\|l_i^{\mathrm{T}} - b_i - A_i(u - u_0)\|_1 + \mathcal{R}(u)\,, \tag{14}$$

where l_i denotes the i^{th} row of L, $u = \mathrm{vec}(\tilde{u})$ and $u_0 = \mathrm{vec}(\tilde{u}_0)$. To obtain a reliable solution we use the well justified assumption, that the disparity map u should be piecewise smooth. We model this assumption by defining $\mathcal{R}(u)$ to be the Total Generalized Variation (TGV) [5], which is a generalization of the well known Total Variation (TV). To be more specific, TGV of second order (TGV^2) will be our choice for the regularization term. Note, that TGV^2 favors piecewise linear solutions, whereas *e.g*. TV favors piecewise constant solutions. This means that the regularization term can be defined as follows

$$\mathcal{R}(u) := \min_{w}\ \alpha_1\|\nabla u - w\|_{\mathcal{M}} + \alpha_0\|\boldsymbol{\nabla} w\|_{\mathcal{M}}\,, \tag{15}$$

where $\|.\|_{\mathcal{M}}$ denotes a Radon norm for vector-valued and matrix-valued Radon measures, and α_0, $\alpha_1 > 0$ are weighting parameters. Also note, that ∇ and $\boldsymbol{\nabla}$

Algorithm 1. Primal-Dual Algorithm for the RPCA Light Field Depth Model

Require: Choose $\sigma > 0$ and $\tau > 0$, s.t. $\tau\sigma = 1$. Set $\Sigma_{p_u}^{-1}$, $\Sigma_{p_w}^{-1}$, T_u^{-1}, and T_w^{-1} as explained in the text, $n = 0$, and the rest arbitrary.

while $n < iter$ **do**

// Dual step
$p_u^{n+1} \leftarrow \mathcal{P}_{\{\|\cdot\|_\infty \leqslant 1\}} \left(p_u^n + \sigma \Sigma_{p_u}^{-1} \alpha_1 \left(\nabla \bar{u}^n - \bar{w}^n\right)\right)$
$p_w^{n+1} \leftarrow \mathcal{P}_{\{\|\cdot\|_\infty \leqslant 1\}} \left(p_w^n + \sigma \Sigma_{p_w}^{-1} \alpha_0 \left(\boldsymbol{\nabla} \bar{w}^n\right)\right)$
for $1 \leqslant i \leqslant M$ **do**
$p_i^{n+1} \leftarrow \mathcal{P}_{\{\|\cdot\|_\infty \leqslant 1\}} \left(p_i^n + \frac{\sigma}{2} \lambda \left(\boldsymbol{D}\boldsymbol{B}(\bar{l}_i^n)^{\mathrm{T}} - b_i - A_i(\bar{u}^n - u_0)\right)\right)$
end for

// Primal step
$u^{n+1} \leftarrow u^n - \tau \, \mathrm{T}_u^{-1} \left(\alpha_1 \nabla^{\mathrm{T}} p_u^{n+1} - \lambda \sum_i A_i p_i^{n+1}\right)$
$w^{n+1} \leftarrow w^n - \tau \, \mathrm{T}_w^{-1} \left(\alpha_0 \boldsymbol{\nabla}^{\mathrm{T}} p_w^{n+1} - \alpha_1 p_u^{n+1}\right)$
for $1 \leqslant i \leqslant M$ **do**
$l_i^{n+1} \leftarrow l_i^n - \frac{\tau}{\lambda^2} \lambda \boldsymbol{B}^{\mathrm{T}} \boldsymbol{D}^{\mathrm{T}} p_i^{n+1}$
end for
$L^{n+1} \leftarrow (\mathrm{id} + \frac{\tau\mu}{\lambda^2} \partial G)^{-1} \left(L^{n+1}\right)$

$\bar{u}^{n+1} \leftarrow 2\, u^{n+1} - u^n$
$\bar{w}^{n+1} \leftarrow 2\, w^{n+1} - w^n$
$\bar{L}^{n+1} \leftarrow 2\, L^{n+1} - L^n$

// Iterate
$n \leftarrow n + 1$

end while

denote finite difference operators, where the first one calculates the finite differences in x and y direction, and the second one is defined as $\boldsymbol{\nabla} := \mathrm{diag}(\nabla, \nabla)$.

We further extend problem (14) to simultaneous super-resolution on all low rank sub-aperture images l_i^{T}. Following the work by Unger *et al.* [34] we introduce linear operators for downsampling and blurring, denoted as $\boldsymbol{D}$ and $\boldsymbol{B}$, respectively.

$$\underset{u,L}{\text{minimize}} \quad \mu \|L\|_* + \lambda \sum_{i=1}^{M} \|\boldsymbol{D}\boldsymbol{B} l_i^{\mathrm{T}} - b_i - A_i(u - u_0)\|_1 + \mathcal{R}(u)\,, \tag{16}$$

where the low rank component L is now computed at a higher resolution.

3 Optimization

In this section we describe how to optimize the proposed RPCA light field model (16). We start with reformulating the problem into a saddle-point

formulation. Therefore we introduce the dual variables p_u, p_w, and p_i ($1 \leqslant i \leqslant M$) and obtain the following formulation

$$\min_{u,w,L} \max_{\substack{\|p_u\|_\infty \leqslant 1 \\ \|p_w\|_\infty \leqslant 1 \\ \|p_i\|_\infty \leqslant 1}} \Big\{ \mu \|L\|_* + \lambda \sum_{i=1}^{M} \langle \boldsymbol{DB} l_i^{\mathrm{T}} - b_i - A_i(u - u_0), p_i \rangle + \alpha_1 \langle \nabla u - w, p_u \rangle + \alpha_0 \langle \boldsymbol{\nabla} w, p_w \rangle \Big\}, \tag{17}$$

where $\langle ., . \rangle$ denotes the standard inner product. This problem can be further rewritten into the following standard form

$$\min_{\hat{x} \in X} \max_{\hat{y} \in Y} \langle K\hat{x}, \hat{y} \rangle + G(\hat{x}) - F^*(\hat{y}), \tag{18}$$

with

$$K = \begin{bmatrix} \alpha_1 \nabla & -\alpha_1 \,\mathrm{id} & 0 & \dots & 0 \\ 0 & \alpha_0 \boldsymbol{\nabla} & 0 & \dots & 0 \\ -\lambda A_1 & 0 & \lambda \boldsymbol{DB} & \dots & 0 \\ \vdots & \vdots & & \ddots & \\ -\lambda A_M & 0 & 0 & & \lambda \boldsymbol{DB} \end{bmatrix}, \quad \hat{x} = \begin{bmatrix} u \\ w \\ l_1^T \\ \vdots \\ l_M^T \end{bmatrix}, \quad \hat{y} = \begin{bmatrix} p_u \\ p_w \\ p_1 \\ \vdots \\ p_M \end{bmatrix}, \tag{19}$$

where id denotes the identity operator. Furthermore, $G(\hat{x}) = \mu \|L\|_*$ and $F^*(\hat{y})$ contains the remaining terms in (17). Also note that $G(\hat{x})$ only operates on the variable L, thus we will redefine it to consider only those variables, *i.e.* $G(L) = \|L\|_*$. A problem of the form (18) can then be solved using the first-order primal dual algorithm proposed by Chambolle *et al.* [9]. Furthermore, we also use positive-definite preconditioning matrices Σ and T to improve the convergence speed of the algorithm as proposed by Pock *et al.* [29]. Here T represents a diagonal matrix of the same size as K, where each diagonal element represents the squared ℓ^2 norm of the corresponding column of K. Σ is calculated in a similar way, but now each diagonal element represents the ℓ^0 norm of the corresponding row in K. The final update scheme is shown in Algorithm 1, where Σ_{p_u} and Σ_{p_w} represent blockdiagonal matrices of Σ that correspond to the dual variables p_u and p_w, respectively. Likewise T_u and T_w represent the according blockdiagonal matrices of T for u and w, respectively. Further, $\mathcal{P}_{\{\|.\|_\infty \leqslant 1\}}$ denotes the reprojection operator, w.r.t. the ℓ^∞ norm denoted as $\|.\|_\infty$, and $(\mathrm{id} + \tau \partial G)^{-1}(L)$ is the proximity operator of the function $G(L)$, which can be calculated by minimizing the following problem.

$$(\mathrm{id} + \tau \partial G)^{-1}(L) = \operatorname*{argmin}_X \left(\|X\|_* + \frac{1}{2\tau} \|X - L\|_F^2 \right) \tag{20}$$

Problem (20) can be solved using spectral soft thresholding. In order to do so we first calculate the singular value decomposition (SVD) of L, *i.e.*

$$L = U \tilde{\Sigma} V^{\mathrm{T}}, \quad \text{with} \quad \tilde{\Sigma} = \mathrm{diag}\left(\sigma_1(L), \dots, \sigma_{\mathrm{rank}(L)}(L)\right), \tag{21}$$

and then apply the soft thresholding operation on each singular value, which yields

$$(\mathrm{id} + \tau \partial G)^{-1}(L) = U \operatorname{diag}\left((\sigma_1(L) - \tau)_+ , \ldots, \left(\sigma_{\operatorname{rank}(L)}(L) - \tau\right)_+ \right) V^{\mathrm{T}}, \quad (22)$$

where $(x)_+ := \max\{0, x\}$.

This concludes the optimization scheme, which solves problem (14). In a final step, we embed Algorithm 1 into a coarse to fine warping scheme [6], which is necessary because of the linearization involved in (11). Moreover, it is also worth to mention, that one can use a structure texture decomposition [39] on the input images to cope with illumination changes.

4 Experimental Results

In this section we will evaluate the proposed algorithm on synthetic and real world scenes. For the synthetic evaluation we use the Light Field Benchmark Dataset (LFBD) [38]. This dataset contains synthetically generated light fields, where each light field is represented by 81 sub-aperture images arranged on a regular 9×9 grid. The light fields are rendered using Blender[1], and the dataset additionally provides a ground truth depth for each sub-aperture image.

We also present a qualitative real world evaluation for light fields from the Stanford Light Field Archive. The light fields in this dataset are captured using a multi-camera array [40] and contain 289 views on a 17×17 grid. Moreover, we also present some qualitative real-world results for light fields captured with the consumer Lytro camera[2]. The Lytro camera captures light fields with a spatial resolution of 380×330 microlenses and a directional resolution of 10×10 pixels per microlens.

4.1 Synthetic Evaluation

We start with the synthetic evaluation. Here we compare our approach to the work by Heber *et al.* [16]. They proposed a variational model with a pointwise ℓ^1 data term combined with an image driven TGV regularization term (ITGV), *i.e.* the prior is connected with the image content via an anisotropic diffusion tensor. Moreover, their model selects a predefined reference view (in this case this is the center view) and matches all the other views against the reference view. Thus, contrary to the proposed model, it only uses a subset of all possible matching combinations between the different views, and the views are also not equal important, *i.e.* the model encodes basically a one vs. all matching.

For the experiments we define the warping center to be the center view of the light field and we extract sub-aperture images with a predefined baseline to the warping center. More precisely, we set $\varphi_1 = \mathbf{0}$ (center view), and define the vectors φ_i for $2 \leqslant i \leqslant 9$ in (10) such that they all have the same length R,

[1] `http://www.blender.org/`
[2] `https://www.lytro.com/`

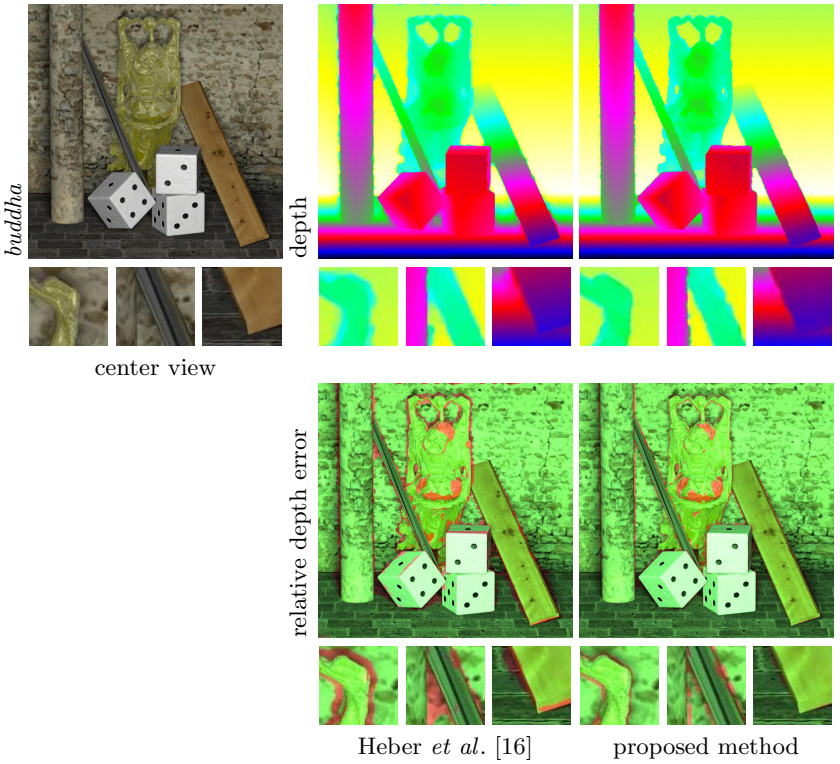

Fig. 2. Qualitative and quantitative results for the *buddha* scene of the Light Field Benchmark Dataset (LFBD). The figure shows the center view of the light field, followed by the color coded depth maps and error maps for the method proposed by Heber *et al.* [16] as well as for the proposed method. The error maps show in green (red) the pixels with a relative depth error of less (more) than 0.2%. Note that the proposed method is much more accurate, especially at occlusion boundaries, due to the robust all vs. all matching term.

and such that their directions are evenly distributed. Also note, that we use the same baseline R as in [16], which results in the same experimental setting as in [16], and thus allows to draw a comparison. Also note, that we do not use the extension for super-resoluation in this case, *i.e.* $\boldsymbol{D} = \boldsymbol{B} = \mathrm{id}$.

Figure 2 shows an example depth map result for the method proposed in [16] and for the proposed RPCA light field model. By considering the closeup views of the depth map results, one sees that the proposed method achieves a higher accuracy, especially at occlusion boundaries. Figure 2 also presents a comparisons in terms of the relative depth error. We highlighted the regions with a relative depth error larger (smaller) than 0.2% in red (green). Note, that an evaluation based on a smaller relative depth error than 0.2% is not meaningful on this daterset, due to the fact that the depth discretization of the provided

Table 1. Quantitative results for the Light Field Benchmark Dataset (LFBD). The table shows the percentage of pixels with a relative depth error of more than 0.2% for the different synthetic scenes. Note, that the results for the method proposed in [16] are taken from the according paper. The results for the method proposed by Wanner *et al.* [36] are obtained by running the accompanying source-code.

	buddha	*buddha2*	*mona*	*papillon*	*stillLife*	*horses*	*medieval*
Wanner *et al.* [36]	7.28	26.55	15.08	16.64	4.50	16.44	24.33
Heber *et al.* [16]	8.37	15.05	12.90	8.79	6.33	16.83	11.09
proposed model	**5.03**	**11.52**	**12.75**	**8.00**	**4.20**	**11.78**	**11.09**

ground truth is too low. We again observe that the proposed method is more robust to certain outliers, *e.g.* due to occlusion or specularity. In the case of the buddha scene shown in Figure 2 the proposed model provides a solution, where only 5.03% of the pixels have a relative depth error larger than 0.2%, whereas the one vs. all data-fidelity term used in the method by Heber *et al.* [16] creates a solution with a significantly larger error region of 8.37%. A similar behavior can be observed for the other scenes in the dataset, as can be seen in Table 1. Furthermore, Table 1 also shows results for the method proposed by Wanner *et al.* [36], which calculates a globally consistent depth labeling. Note that this comparison might not be very fair, because the results by Wanner *et al.* are obtained by performing a complete grid-search to find the best parameter settings, whereas the other methods are only hand-tuned. However, the results show that the method proposed in [16] already outperforms the method by Wanner *et al.* [36] on several scenes by quite a bit. Finally, the proposed model outperforms both competitors on the complete dataset, but the better performance comes at the price of a higher computational time of several minutes. It is also worth to mention, that the method proposed by Tao *et al.* [32] fails on this dataset completely, as also reported in their paper.

Super-resolution

Next we present super-resolution results for the extended RPCA light field model (16). We define the downsampling operator $\boldsymbol{D}$ and the blurring operator $\boldsymbol{B}$ as proposed by Unger *et al.* [34]. Figure 3 shows closeup views of the obtained upsampling results for two scenes of the LFBD. Here we used 21 sub-aperture images for the reconstruction, where the low rank components of the warped sub-aperture images have been magnified by a factor of three. By considering the super-resolved results shown in Figure 3 one recognizes a clear increase in sharpness.

4.2 Real World Experiments

Now we continue with the real-world evaluation. Figure 4 present a qualitative comparison to the method proposed by Wanner *et al.* [36]. Here we use a light

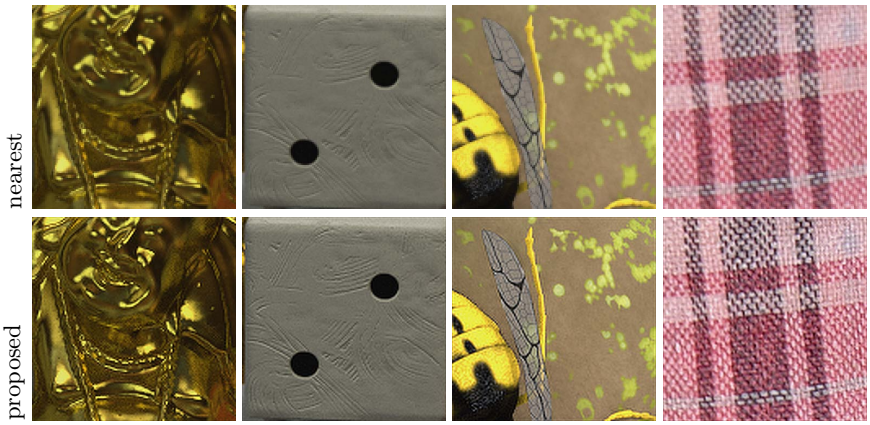

Fig. 3. Qualitative results of the extended version of the RPCA light field model (*cf.* (16)) for the *buddha2* and *stillLife* scene of the Light Field Benchmark Dataset (LFBD). The figure shows closeup views of the nearest neighbor interpolated center view, as well as closeup views of one super-resolved low rank component, where the super-resolved results provide increased sharpness.

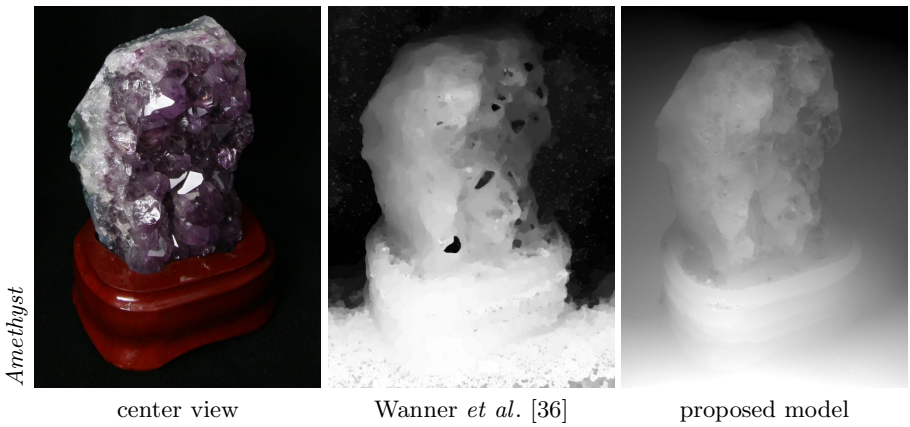

Fig. 4. Qualitative comparison for a light field from the Stanford Light Field Archive. The figure shows from left to right, the center view of the light field, the results for the method proposed by Wanner *et al.* [36] (image is taken from their paper) and the result for the proposed RPCA light field model.

Fig. 5. Qualitative results, for light fields captured with a Lytro camera. The figure shows color coded disparity maps as well as the according center views of the light field.

field from the Stanford Light Field Archive as input for our algorithm, where we extract 17 sub-aperture images with evenly spread directional offsets φ_i ($1 \leqslant i \leqslant 17$). Also note, that the scene is quite challenging, due to reflective and specular surface. By comparing the results one sees that the proposed method allows to create a solution with significantly more details and fewer outliers, by approximately the same amount of regularization. The reason is on one side the continuous formulation of the proposed model, and on the other side the robust implicit all vs. all matching term.

In Figure 5 we also present results for light fields captured with the Lytro camera. Therefore, we extract 17 sub-aperture images from the raw images captured with such a camera. Note, that these sub-aperture images have a quite low resolution of 380×330, and include a significant amount of noise and outliers. Nevertheless, the proposed method is capable to create piecewise smooth depth maps, with clear depth discontinuities. Also note, that the proposed method performs particularly well in this case, due to the implicit denoising of the warped views.

5 Conclusion

In this paper we proposed a global matching term for a multi-view stereo model, which has not been considered before for this task. We formulated our model to perform a low rank minimization on the stack of warped images, which can also be interpreted as an all vs. all matching between the images in the stack. We showed how to relax the according combinatorial problem to a convex optimization problem, by using a nuclear norm and ℓ^1 norm relaxation. The proposed variational model assumes a certain amount of overlap in the warped views. Thus we tested it on light field data, where this assumption is obviously fulfilled. We also want to point out, that the proposed RPCA matching term is not limited to the light field setting. In general such a matching term is well suited for all kind of problems with highly redundant input data.

Finally we want to mention, that the proposed model can still be further refined by performing additionally iterative ℓ^1 reweighting. Such a refinement

procedure can be applied on the ℓ^1 term in problem (14) [8], as well as on the nuclear norm [27], to further increase the accuracy especially at depth discontinuities. Implementing and evaluating such a refinement is left as future work.

References

1. Adelson, E.H., Wang, J.Y.A.: Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 99–106 (1992)
2. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Computation 15, 1373–1396 (2002)
3. Bishop, T.E., Favaro, P.: The light field camera: Extended depth of field, aliasing, and superresolution. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(5), 972–986 (2012)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. 23(11), 1222–1239 (2001)
5. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences 3(3), 492–526 (2010)
6. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 25–36. Springer, Heidelberg (2004)
7. Candès, E.J., Li, X., Ma, Y., Wright, J.: Robust principal component analysis? J. ACM 58(3), 1–37 (2011)
8. Candès, E.J., Wakin, M.B., Boyd, S.P.: Enhancing sparsity by reweighted ℓ_1 minimization (2007)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision 40, 120–145 (2011)
10. Chen, S.S., Donoho, D.L., Saunders, M.A.: Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing 20, 33–61 (1998)
11. Coffey, D.F.W.: Apparatus for making a composite stereograph (December 1936)
12. Collins, R.T., Collins, R.T.: A space-sweep approach to true multi-image matching (1996)
13. Dudnikov, Y.A.: Autostereoscopy and integral photography. Optical Technology 37(3), 422–426 (1970)
14. Eckart, C., Young, G.: The approximation of one matrix by another of lower rank. Psychometrika 1, 211–218 (1936)
15. Fife, K., Gamal, A.E., Philip Wong, H.S.: A 3mpixel multi-aperture image sensor with 0.7m pixels in 0.11m cmos (February 2008)
16. Heber, S., Ranftl, R., Pock, T.: Variational Shape from Light Field. In: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (2013)
17. Hotelling, H.: Analysis of a complex of statistical variables into principal components. J. Educ. Psych. 24 (1933)
18. Isaksen, A., McMillan, L., Gortler, S.J.: Dynamically reparameterized light fields. In: SIGGRAPH, pp. 297–306 (2000)
19. Jolliffe, I.T.: Principal Component Analysis. Springer, Berlin (1986)

20. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. ACM Trans. Graph. 26(3) (July 2007)
21. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, pp. 31–42. ACM, New York (1996)
22. Lippmann, R.: La photographie intégrale. Comptes-Rendus, Académie des Sciences 146, 446–551 (1908)
23. Ng, R.: Digital Light Field Photography. Phd thesis, Stanford University (2006), http://www.lytro.com/renng-thesis.pdf
24. Ng, R., Levoy, M., Brédif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Tech. rep., Stanford University (2005)
25. Okutomi, M., Kanade, T.: A multiple-baseline stereo. IEEE Trans. Pattern Anal. Mach. Intell. 15(4), 353–363 (1993)
26. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(11), 2233–2246 (2012)
27. Peng, Y., Suo, J., Dai, Q., Xu, W., Lu, S.: Robust image restoration via reweighted low-rank matrix recovery. In: Gurrin, C., Hopfgartner, F., Hurst, W., Johansen, H., Lee, H., O'Connor, N. (eds.) MMM 2014, Part I. LNCS, vol. 8325, pp. 315–326. Springer, Heidelberg (2014)
28. Perwass, C., Wietzke, L.: Single lens 3d-camera with extended depth-of-field (2012)
29. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: International Conference on Computer Vision (ICCV), pp. 1762–1769. IEEE (2011)
30. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision 47(1-3), 7–42 (2002)
31. Stühmer, J., Gumhold, S., Cremers, D.: Real-time dense geometry from a handheld camera. In: Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) Pattern Recognition. LNCS, vol. 6376, pp. 11–20. Springer, Heidelberg (2010)
32. Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R.: Depth from combining defocus and correspondence using light-field cameras (December 2013)
33. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science 290(5500), 2319–2323 (2000)
34. Unger, M., Pock, T., Werlberger, M., Bischof, H.: A convex approach for variational super-resolution. In: Goesele, M., Roth, S., Kuijper, A., Schiele, B., Schindler, K. (eds.) Pattern Recognition. LNCS, vol. 6376, pp. 313–322. Springer, Heidelberg (2010)
35. Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans. Graph. 26(3) (July 2007)
36. Wanner, S., Goldluecke, B.: Globally consistent depth labeling of 4D lightfields. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012)
37. Wanner, S., Goldluecke, B.: Spatial and angular variational super-resolution of 4D light fields. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part V. LNCS, vol. 7576, pp. 608–621. Springer, Heidelberg (2012)
38. Wanner, S., Meister, S., Goldluecke, B.: Datasets and benchmarks for densely sampled 4D light fields. In: Vision, Modelling and Visualization (VMV) (2013)

39. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An Improved Algorithm for TV-L^1 Optical Flow. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Statistical and Geometrical Approaches to Visual Motion Analysis. LNCS, vol. 5604, pp. 23–45. Springer, Heidelberg (2009)
40. Wilburn, B., Joshi, N., Vaish, V., Talvala, E.V., Antunez, E., Barth, A., Adams, A., Horowitz, M., Levoy, M.: High performance imaging using large camera arrays. ACM Trans. Graph. 24(3), 765–776 (2005)