Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks

Reuben Farrugia, Christine Guillemot

Light Field Super-Resolution using a Low-Rank Prior and Deep Convolutional Neural Networks

Reuben A. Farrugia, *Senior Member, IEEE,* Christine Guillemot, *Fellow, IEEE,*

Abstract—Light field imaging has recently known a regain of interest due to the availability of practical light field capturing systems that offer a wide range of applications in the field of computer vision. However, capturing high-resolution light fields remains technologically challenging since the increase in angular resolution is often accompanied by a significant reduction in spatial resolution. This paper describes a learning-based spatial light field super-resolution method that allows the restoration of the entire light field with consistency across all angular views. The algorithm first uses optical flow to align the light field and then reduces its angular dimension using low-rank approximation. We then consider the linearly independent columns of the resulting low-rank model as an embedding, which is restored using a deep convolutional neural network (DCNN). The super-resolved embedding is then used to reconstruct the remaining views. The original disparities are restored using inverse warping where missing pixels are approximated using a novel light field inpainting algorithm. Experimental results show that the proposed method outperforms existing light field super-resolution algorithms, achieving PSNR gains of 0.23 dB over the second best performing method. The performance is shown to be further improved using iterative back-projection as a post-processing step.

Index Terms—Deep Convolutional Neural Networks, Light Field, Low-Rank Matrix Approximation, Super-Resolution.

1 Introduction

Light field imaging has emerged as a promising technology for a variety of applications going from photo-realistic image-based rendering to computer vision applications such as 3D modeling, object detection, classification and recognition. As opposed to traditional photography which captures a 2D projection of the light in the scene, light fields collect the radiance of light rays along different directions [1], [2]. This rich visual description of the scene offers powerful capabilities for scene understanding and for improving the performance of traditional computer vision problems such as depth sensing, post-capture refocusing, segmentation, video stabilization and material classification to mention a few.

Light fields acquisition devices have been recently designed, going from rigs of cameras [2] capturing the scene from slightly different viewpoints to plenoptic cameras using micro-lens arrays placed in front of the photo-sensor [1]. These acquisition devices offer different trade-offs between angular and spatial resolution. Rigs of cameras capture views with a high spatial resolution but in general with a limited angular sampling hence large disparities between views. On the other hand, plenoptic cameras capture views with a high angular sampling, but at the expense of a limited spatial resolution. In plenoptic cameras, the angular sampling is related to the number of sensor pixels located behind each microlens, while the spatial sampling is related to the number of microlenses.

- *R.A. Farrugia is with the Department of Communications and Computer Engineering, University of Malta, Malta, e-mail: (reuben.farrugia@um.edu.mt).*
- *C. Guillemot is with the Institut National de Recherche en Informatique et en Automatique, Rennes 35042, France, e-mail: (christine.guillemot@intria.fr).*

Manuscript received January, 2018; revised xxx.

Light fields represent very large volumes of high dimensional data bringing new challenges in terms of capture, compression, editing and display. The design of efficient light field image processing algorithms, going from analysis, compression to super-resolution and editing has thus recently attracted interest from the research community. A comprehensive overview of light field image processing techniques can be found in [3].

This paper addresses the problem of light field spatial super-resolution. Single image super-resolution has been an active field of research in the past years, leading to quite mature solutions. However, super-resolving each view separately using state of the art single-image super-resolution techniques would not take advantage of light field properties, in particular of angular redundancy which depends on scene geometry [4]. Moreover, considering each view as a separate entity may reconstruct light fields which are angularly incoherent [5]. Driven by this observation, several light field super-resolution methods have been proposed that try to increase the spatial resolution of the light field by exploiting its geometrical structure [5], [6], [7], [8], [9], [10], [11], [12].

In this paper, we propose a spatial light field super-resolution method using a deep CNN (DCNN) with ten convolutional layers. Instead of using DCNN to restore each view independently, as done in [10], [11], we restore all angular views within a light field simultaneously. This allows us to exploit both spatial and angular information and thus generate light fields which are angularly coherent. A Naïve approach would be to train a DCNN with $n = P \times Q$ inputs, where P and Q represent the number of vertical and horizontal angular views respectively. However, this would significantly increase the complexity of the DCNN, making it harder to train and more prone to over-fitting (see results in Figure 5 and discussions in Section X). Instead,

given that each angular view captures the same scene from a different view point, we align all angular views to the centre view using optical flow and then reduce the angular dimension of the aligned light field using a low-rank model of rank k, where $k \ll n$. Results in section 4.1 show that the alignment allows us, with the considered low rank model, to significantly reduce the angular dimension of the light field. The linearly independent column-vectors of the low-rank representation of the aligned light field, which constitute an embedding of the light field views in a lower-dimensional space, are then considered as a volume and simultaneously restored using a DCNN with k input channels. This allows us to significantly reduce the complexity of the network which is easier to train while still preserving angular consistency. The restored column-vectors are then combined to reconstruct the aligned high-resolution light field. In the final stage we use inverse warping to restore the original disparities of the light field and fill the cracks caused by occlusion using a novel diffusion based inpainting strategy that propagates the restored pixels along the dominant orientation of the EPI.

Simulation results demonstrate that the proposed method outperforms all other schemes considered here when tested on 13 different light fields from two different datasets. It is important to mention that our method was not trained on the Stanford light fields and the results in section 5 clearly show that our proposed method generalizes well even when considering light field structures whose disparities are significantly larger than those used for training. Further analysis in section 5 shows that additional gain in performance can be achieved using iterative back projection (IBP) as a post processing step. These results show that our method can significantly outperform existing light field super-resolution methods including the deep learning-based light field super-resolution method presented in [11].

The paper is organized as follows. Section 2 presents the related work while the notations used in the paper are introduced in section 3. The proposed method is described in section 4. Section 5 discusses the simulation results with different types of light fields and provide the final concluding remarks in section 6.

2 Related Work

2.1 Light Field Super-Resolution

Assuming that the low-resolution light field captures the scene from a different viewing angle, the problem can be posed as the one of recovering the high-resolution (HR) views from multiple low- resolution images with unknown non integer displacements. A number of methods hence proceed in two steps. A first step consists in estimating the disparities using depth or disparity estimation techniques. The HR light field views are then found using Bayesian or variational optimization frameworks with different priors. This is the case in [6] and [7] where the authors first recover a depth map and formulate the spatial light field super-resolution problem either as a simple linear problem [6] or as a Bayesian inference problem [7] assuming an image formation model with Lambertian reflectance priors and a depth-dependent blurring kernel. A Gaussian mixture model (GMM) is proposed instead in [8] to address denoising, spatial and angular super-resolution of light fields. The reconstructed 4D-patches are estimated using a linear minimum mean square error (LMMSE) estimator, assuming a disparity-dependent GMM for the patch structure. In [9], the geometry is estimated by computing structure tensors in the Epipolar Plane Images (EPI). A variational optimization framework is then used to spatially super-resolve the different views given their estimated depth maps and to increase the angular resolution.

Another category of methods is based on machine learning techniques which learn a model of correspondences between low- and high-resolution data. In [5], the authors learn projections between low-dimensional subspaces of 3D patch-volumes of low- and high-resolution, using ridge regression. Data-driven learning methods based on deep neural network models have been recently shown to be quite promising for light fields super-resolution. In [10], stacked input images are up-scaled to a target resolution using bicubic interpolation and super-resolved using a spatial convolutional neural network (CNN). This spatial CNN learns a non-linear mapping between low- and high-resolution views. Its output is then fed into a second CNN to perform angular super-resolution. The approach in [10] takes, at the input of the spatial CNN, pairs or 4-tuples of neighboring views, leading to three spatial CNNs to be learned. A single CNN is proposed by the same authors in [11] to process each view independently. A shallow neural network is proposed in [12] to restore light fields captured by a plenoptic camera. Each lenslet micro-image of size $A \times A$, containing pixels corresponding to the same 3D point seen from different views, is fed into an angular neural network which actually spatially super-resolve the input lenslet region into a $2A \times 2A$ micro-image. A second neural network then processes the resulting micro-images, per groups of four, to generate three novel pixels corresponding to a magnification factor of $\times 2$ horizontally and vertically. The method, while suitable for a magnification factor of $\times 2$, cannot be easily extended to other magnification factors.

The problem of angular super-resolution of light fields is also addressed in [13] using an architecture based on two CNNs, one CNN being used to estimate disparity maps and the second CNN being used to synthesis intermediate views. The authors in [14] define a CNN architecture in the EPI to increase the angular resolution.

Hybrid imaging systems camera have also been considered to overcome the fixed spatial and angular resolution trade-off of plenoptic cameras [15], [16]. In [15], the HR image captured by the DSLR camera is used to super-resolve the low-resolution images captured by an Illum light field camera. The authors in [16] describe an acquisition device formed by eight low-resolution side cameras arranged around a central high-quality SLR lens. A super-resolution method, called iterative patch- and depth-based synthesis (iPADS), is then proposed to reconstruct a light field with the spatial resolution of the SLR camera and an increased number of views.

2.2 Low-Rank Approximation

Singular value decomposition (SVD) is a classical technique that can be used to approximate a matrix by a low-rank

model. Robust principal component analysis (RPCA) was introduced in [17] to decompose a matrix as a sum of a low-rank and a sparse error matrix. RPCA was extended in [18] to search for the homographies that globally align a batch of linearly correlated images. More recently, the authors in [19] tried to reduce the redundancy present in light fields by jointly aligning the angular views in the light field and estimating a low-rank approximation model of the light field. Both methods [18], [19] seek for an optimal set of homographies such that the matrix of aligned images can be decomposed in a low-rank matrix of aligned images, with the former constraining the error matrix to be sparse.

2.3 Light Field Inpainting

Light field inpainting involves the editing of the centre view followed by the propagation of the restored information to all the other views. A 3D voxel-based model of the scene with associated radiance function was proposed in [20] to propagate the edited information from the center view of a light field to all the other views. A method based on reparametrization of the light field was proposed in [21] while the depth information was used in [22] to propagate the edits from the center view to all the other views while preserving the angular coherence. The authors in [23] use tensor driven diffusion to propagate information along the Epipolar Plane Image (EPI) structure of the light field.

3 Notation and Problem Formulation

We consider here the simplified 4D representation of light fields called 4D light field in [24] and lumigraph in [25], describing the radiance along rays by a function $I(x, y, s, t)$ where the pairs (x, y) and (s, t) respectively represent spatial and angular coordinates. The light field can be seen as capturing an array of viewpoints of the scene with varying angular coordinates (s, t). The different views will be denoted here by $\mathbf{I}_{s,t} \in \mathbb{R}^{X,Y}$, where X and Y represent the vertical and horizontal dimension of each view.

In the following, the notation $\mathbf{I}_{s,t}$ for the different views will be simplified as $\mathbf{I}_i$ with a bijection between (s, t) and i. The complete light field can hence be represented by a matrix $\mathbf{I} \in \mathbb{R}^{m,n}$:

$$\mathbf{I} = [vec(\mathbf{I}_1) \mid vec(\mathbf{I}_2) \mid \cdots \mid vec(\mathbf{I}_n)] \tag{1}$$

with $vec(\mathbf{I}_i)$ being the vectorized representation of the $i-$th angular view, m represents the number of pixels in each view ($m = X \times Y$) and n is the number of views in the light field ($n = P \times Q$), where P and Q represent the number of vertical and horizontal angular views respectively.

Let $\mathbf{I}^H$ and $\mathbf{I}^L$ denote the high- and low- resolution light fields, respectively. The super-resolution problem can be formulated in Banach space as

$$\mathbf{I}^L = \downarrow_\alpha \mathbf{B}\mathbf{I}^H + \boldsymbol{\eta} \tag{2}$$

where $\boldsymbol{\eta}$ is an additive noise matrix, $\downarrow_\alpha$ is a downsampling operator applied on each angular view where α is the magnification factor and $\mathbf{B}$ is the blurring kernel. There are many possible high-resolution light fields $\mathbf{I}^H$ which can produce the input low-resolution light field $\mathbf{I}^L$ via the acquisition model defined in (2). Hence, solving this ill-posed inverse problem requires introducing some priors on $\mathbf{I}^H$, which can be a statistical prior such as a GMM model [8], or priors learned from training data as in [5], [10], [11].

Another way to visualize a light field is to consider the EPI representation. An EPI is a spatio-angular slice from the light field, obtained by fixing one of the spatial coordinates and one of the angular coordinates. Consider we fix $y := y*$ and $t := t*$, an EPI is an image defined as $\boldsymbol{\epsilon}_{y*,t*} := \mathbf{I}(x, y*, s, t*)$. Alternatively, the vertical EPI is obtained by fixing $x := x*$ and $s := s*$. Figure 6b shows a typical EPI structure, where the slopes of the isophote lines in the EPI are related to the disparity between the views [9]. Isophote lines with a slope of $\pi/2\,\mathrm{rad}$ indicate that there is no disparity across the views while, the larger is the difference between the slope and $\pi/2\,\mathrm{rad}$, the larger is the disparity across the views.

4 Proposed Method

Figure 1 depicts the block diagram of the proposed spatial light field super-resolution algorithm. Each angular view of the low-resolution light field $\mathbf{I}^L$ is first bicubic interpolated so that both $\mathbf{I}^H$ and $\mathbf{I}^L$ have the same resolution.

The goal here is to restore all the views simultaneously in order to guarantee that the reconstructed views are angularly coherent. However, a light field consists of a very large volume of high-dimensional data, with obvious implications on the complexity of the neural network and on the needed amount of training data. Fortunately, it also contains a lot of redundant information since every angular view captures the same scene from a different viewpoint. Moreover, different light field capturing devices have different spatial and angular specifications, which makes it very hard for a learning-based algorithm to learn a generalized model suitable to restore all kind of light fields irrespective of the capturing device. The *Dimensionality Reduction* module tries to solve both problems simultaneously where it uses optical flow to align the light field and a low-rank matrix approximation to reduce the dimension of the light field. Results in section 4.1 show that we can reduce the dimensionality of the light field from $\mathbb{R}^{m,n}$ to $\mathbb{R}^{m,k}$, where $k \ll n$ is the rank of the matrix, while preserving most of the information contained in the light field.

The *Light Field Restoration* module then considers the k linearly independent column-vectors of the rank-k representation of the low-resolution light field as an embedding of the light field. We then use a DCNN to recover the texture details of the light field embedding in the lower dimensional space. The super-resolved embedding gives an estimate of the aligned high-resolution light field. The *Light Field Reconstruction* module then warps the estimated aligned high-resolution light field to restore the original disparities. Holes corresponding to cracks or occlusions are then filled in by diffusing information in the Epipolar Plane Images (EPI) along directions of isophote lines computed, for the positions of missing pixels, in the EPI of the low-resolution light field. *Iterative back-projection* can be further used as a post-process to refine the super-resolved light field and assure that the restored light field is consistent with

the low-resolution light field. More information about each module is provided in the following subsections.

4.1 Light Field Dimensionality Reduction

If we stack the views as the columns of a large matrix $\mathbf{I}$, the angular dimension of the light field can be reduced by searching for a low rank approximation of the matrix $\mathbf{I}$. In order to minimize the rank of the matrix (ideally rank 1), the views (columns) need to be aligned. Figure 2 shows that while both RASL [18] and LRA [19] methods manage to globally align the angular views, the mean view is still very blurred, indicating that the light field is not suitably aligned. The authors in [5] have used the block matching algorithm (BMA) to align patch volumes. The results in Figure 2 show that BMA manages to align better the angular views, where the average variance across the n views is significantly reduced. This result suggests that local methods can improve the alignment of the angular views, which as we will see in the sequel, will allow us to significantly reduce the dimensionality of the light field.

In this paper, we formulate the light field dimensionality reduction problem as

$$\underset{\mathbf{u},\mathbf{v},\mathbf{A}}{min} ||\Gamma_{\mathbf{u},\mathbf{v}}\left(\mathbf{I}^L\right) - \mathbf{A}||_2^2 \quad \text{s.t.} \quad rank(\mathbf{A}) = k \tag{3}$$

where $\mathbf{u} \in \mathbb{R}^{m,n}$ and $\mathbf{v} \in \mathbb{R}^{m,n}$ are flow vectors that specify the displacement of each pixel needed to align each angular view with the centre view, $\mathbf{A}$ is a rank-k matrix which approximates the aligned light field and $\Gamma_{\mathbf{u},\mathbf{v}}(\cdot)$ is a forward warping operator (which here performs a disparity compensation where the disparities maps $(\mathbf{u},\mathbf{v})$ are estimated with an optical flow estimator). This optimization problem is computationally intractable. Instead, we decompose this problem in two sub-problems: i) use optical flow to find the flow matrices $\mathbf{u}$ and $\mathbf{v}$ that best align each angular view with the centre view and ii) use low-rank approximation to derive the rank-k matrix that minimizes the error with respect to the aligned light field.

4.1.1 Optical Flow

The problem of aligning all the angular views with the centre view can be formulated as

$$\mathbf{I}_j(x,y) = \mathbf{I}_i(x+\mathbf{u}_i, y+\mathbf{v}_i) \quad i \in [1,n]\,(i \neq j) \tag{4}$$

where j corresponds to the index of the centre view, and $(\mathbf{u}_i,\mathbf{v}_i)$ are the flow vectors optimal to align the i-th angular view with the centre view. There are several optical flow algorithms intended to solve this problem [26], [27], [28], [29] where Figure 2 shows the performance of some of these methods. It can be seen that the mean aligned view computed using [26] is generally blurred while those aligned using the methods in [28], [29] generally provide ghosting artefacts at the edges. Moreover, it can be seen that SIFT Flow [27] generally provides very good alignment and manages to attain the smallest variation across the angular views. While the SIFT Flow algorithm will be used in this paper to compute the flow vectors, any other optical flow method can be used.

4.1.2 Low-Rank Approximation

Given that the flow-vectors $(\mathbf{u}_i,\mathbf{v}_i)$ for the i-th angular view are already available, the minimization problem in Eq. (3) can now be reduced to

$$\underset{\mathbf{B^L},\mathbf{C^L}}{min} ||\mathbf{I}_\Gamma^L - \mathbf{B}^L\mathbf{C}^L||_2^2 \quad \text{s.t.} \quad rank(\mathbf{B}^L) = k \tag{5}$$

where $\mathbf{I}_\Gamma^L = \Gamma_{\mathbf{u},\mathbf{v}}\left(\mathbf{I}^L\right)$, $\mathbf{B}^L \in \mathbb{R}^{m,k}$ is a rank-k matrix and $\mathbf{C}^L \in \mathbb{R}^{k,n}$ is the combination weight matrix. These matrices can be found using singular value decomposition (SVD) $\mathbf{I}_\Gamma^L = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{B}^L$ is set as the k first columns of $\mathbf{U}\mathbf{\Sigma}$ and $\mathbf{C}^L$ is set as the k first rows of $\mathbf{V}^T$, so that $\mathbf{B}^L\mathbf{C}^L$ is the closest k-rank approximation of the aligned light field $\mathbf{I}_\Gamma^L$. The error matrix $\mathbf{E}^L$ is the error matrix which is simply computed using $\mathbf{E}^L = \mathbf{I}_\Gamma^L - \mathbf{B}^L\mathbf{C}^L$

Figure 3 depicts the performance of three different dimensionality reduction techniques at different ranks. To measure the dimensionality reduction ability of these methods we compute the root mean square error (RMSE) between the aligned original and the rank-k representation of the aligned light field. It can be seen that the RASL algorithm has the largest distortions at almost all ranks when compared to the other two approaches. On the other hand, it can be seen that HLRA manages to significantly outperform RASL. Nevertheless, it can be clearly observed that the proposed Sift Flow + LRA method gives the best performance, especially at lower ranks, indicating that more information is captured within the low-rank matrix[1]. To emphasize this point we show in figure 3 the principal basis of PCA, HLRA and Sift Flow + LRA. PCA is computed on the light field without disparity estimation and therefore can be considered here as a baseline to show that alignment allows us to get more information in the principal basis. Moreover, it can be seen that the principal basis derived using our Sift Flow + LRA manages to capture more texture detail in the principal basis than the other methods and confirms the benefit that local alignment has on the energy compaction ability of the proposed dimensionality reduction method.[2]

4.2 Light Field Restoration

We consider a low-rank representation of the aligned low-resolution light field $\mathbf{A}^L = \mathbf{B}^L\mathbf{C}^L$, where $\mathbf{A}^L \in \mathbb{R}^{m,n}$ is a rank-k matrix with $k \ll n$. Similarly, $\mathbf{A}^H = \mathbf{B}^H\mathbf{C}^H$ is a rank-k representation of the aligned high-resolution light field. The rank of a matrix is defined as the maximum number of linearly independent column vectors in the matrix. Moreover, the linearly dependent column vectors of a matrix can be reconstructed using a weighted summation of the linearly independent column vectors of the same matrix. This leads us to decompose $\mathbf{A}^L$ in two submatrices: $\breve{\mathbf{A}}^L \in \mathbb{R}^{m,k}$ which groups the linear independent column vectors of $\mathbf{A}^L$ and $\hat{\mathbf{A}}^L \in \mathbb{R}^{m,n-k}$ which groups the linearly dependent column vectors of $\mathbf{A}^L$. In practice, we

1. The performance of HLRA improves at higher ranks. However, the gain is relatively small and may be attributed to different implementation details rather than actual performance gains.

2. Note that the RASL method does decompose the matrix into a combination of basis elements and therefore the principal basis of RASL could not be shown here.

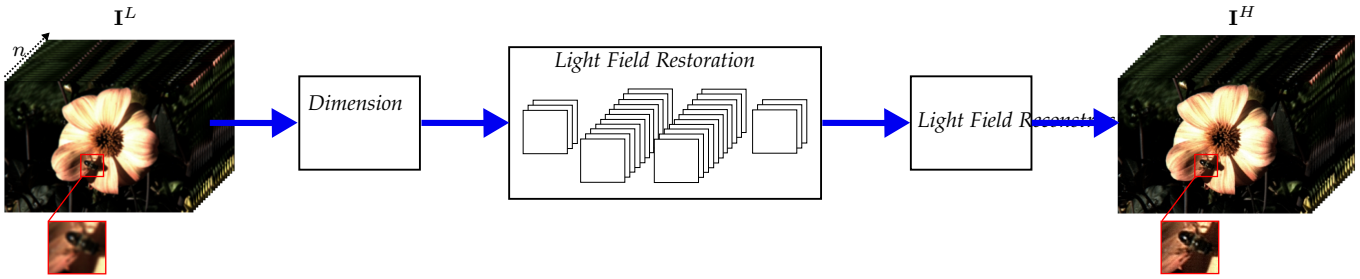


Fig. 1: Block diagram of the proposed light field super-resolution method

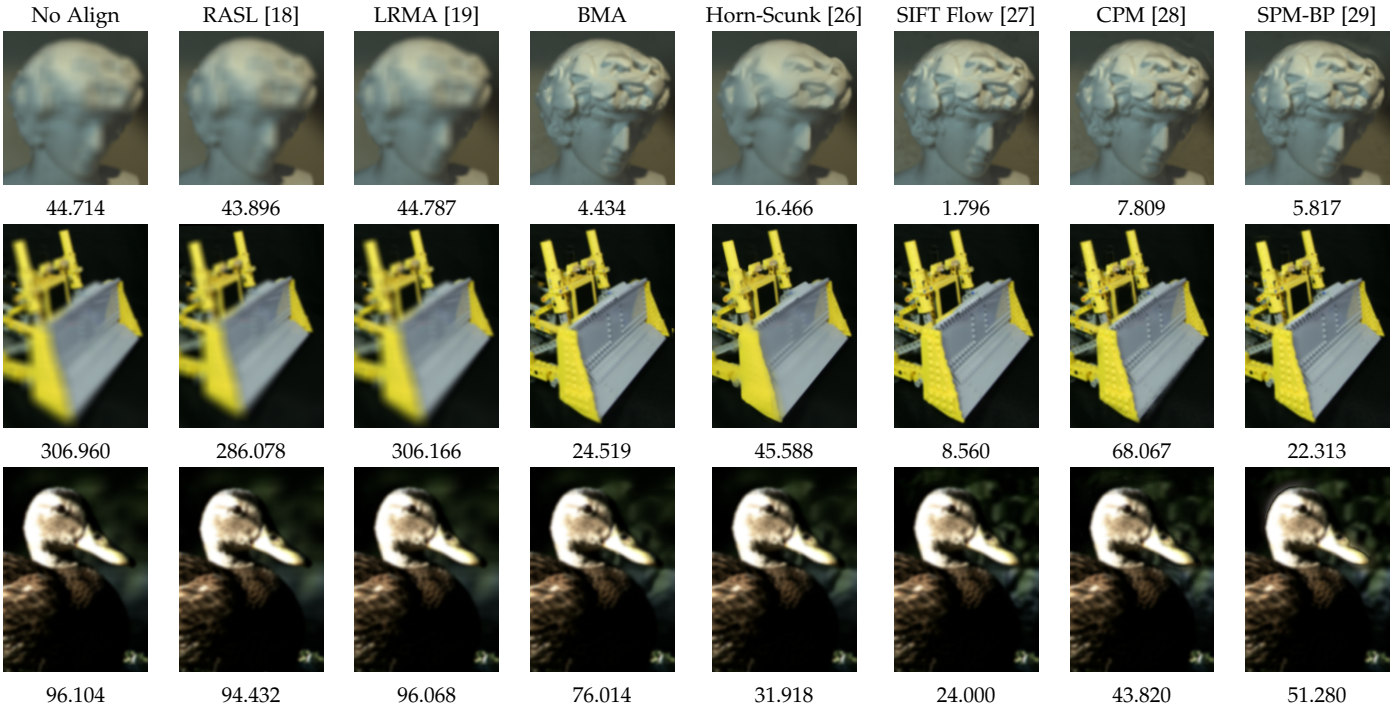


Fig. 2: Cropped regions of the mean angular views when using different disparity compensation methods. Underneath each image we provide the average variance across the n angular views which was used in [5] to characterize the performance of the alignment algorithm, where smaller values indicate better alignment.

decompose the rank-k matrix $\mathbf{A}^L$ using QR decomposition (*i.e.* $\mathbf{A}^L = \mathbf{QR}$). The index of the linearly independent components of $\mathbf{A}^L$ then correspond to the index of the non-zero diagonal elements of the upper-triangular matrix $\mathbf{R}$. We then use the same indices to decompose $\mathbf{A}^H$ into sub-matrices $\breve{\mathbf{A}}^H$ and $\hat{\mathbf{A}}^H$. The matrix $\hat{\mathbf{A}}^L$ can be reconstructed as a linear combination of $\breve{\mathbf{A}}^L$, where the weight matrix $\mathbf{W}$ is computed using

$$\mathbf{W} = \left(\breve{\mathbf{A}}^{L\intercal}\breve{\mathbf{A}}^L\right)^\dagger \breve{\mathbf{A}}^{L\intercal}\hat{\mathbf{A}}^L \tag{6}$$

where $(\cdot)^\dagger$ stands for the pseudo inverse operator. We assume here that the weight matrix $\mathbf{W}$, which is optimal in terms of least squares to reconstruct $\hat{\mathbf{A}}^L$, is suitable to reconstruct $\hat{\mathbf{A}}^H$.

Driven by the recent success of deep learning in the field of single-image [30], [31] and light field super-resolution [10], [11], we use a DCNN to model the upscaling function that minimizes the following objective function

$$1/2||\breve{\mathbf{A}}^H - f\left(\breve{\mathbf{A}}^L\right)||^2 \tag{7}$$

where $f(\cdot)$ is a function modelled by the DCNN illustrated in Figure 4 which has ten convolutional layers. The linearly independent sub-matrix $\breve{\mathbf{A}}^L$ is passed through a stack of convolutional and rectified linear unit (ReLU) layers. We use a convolution stride of 1 pixel with no padding nor spatial pooling. The first convolutional layer has 64 filters of size $3 \times 3 \times k$ while the last layer, which is used to reconstruct the high-resolution light field, employs k filter of size $3 \times 3 \times 64$. All the other layers use 64 filters of size $3 \times 3 \times 64$ which are initialized using the method in [32]. The DCNN was trained using a total of 200,000 random patch-volumes of size $64 \times 64 \times k$ from the 98 low- and high-resolution low-rank approximation of rank k of the light fields from the EPFL, INRIA and HCI datasets[3]. The Titan GTX1080Ti Graphical Processing Unit (GPU) was used to speed up the training process.

During the evaluation phase, we estimate the super-resolved linearly independent representation of the light

3. It must be noted that none of the light fields used for validation were used for training.

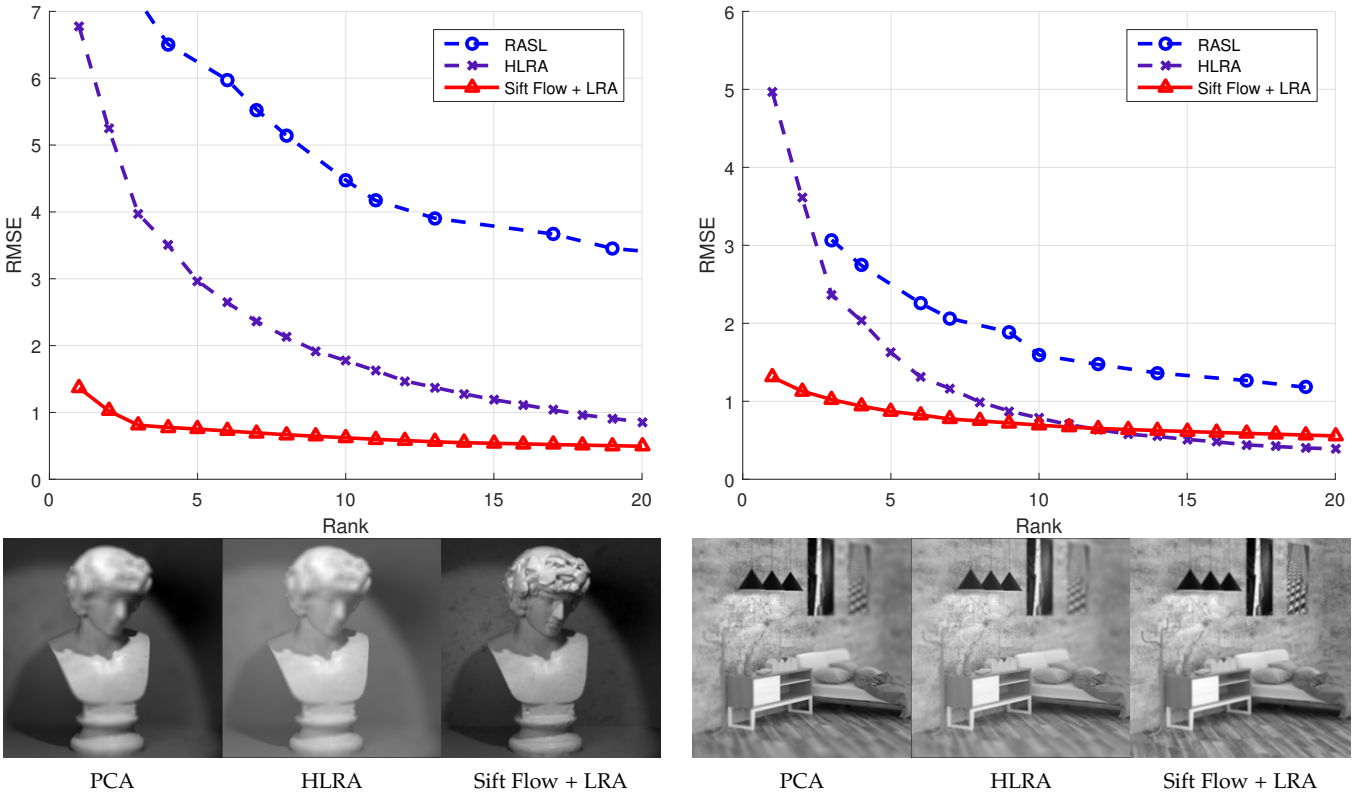


Fig. 3: These figures show how the error between the low-rank and full rank representation vary at different ranks. It can be seen that using optical flow to align the light field followed by low-rank approximation attains the best performance. The images in the second row show the principal basis derived using different methods. The sharper the principal basis is the more information is being captured in the principal basis.

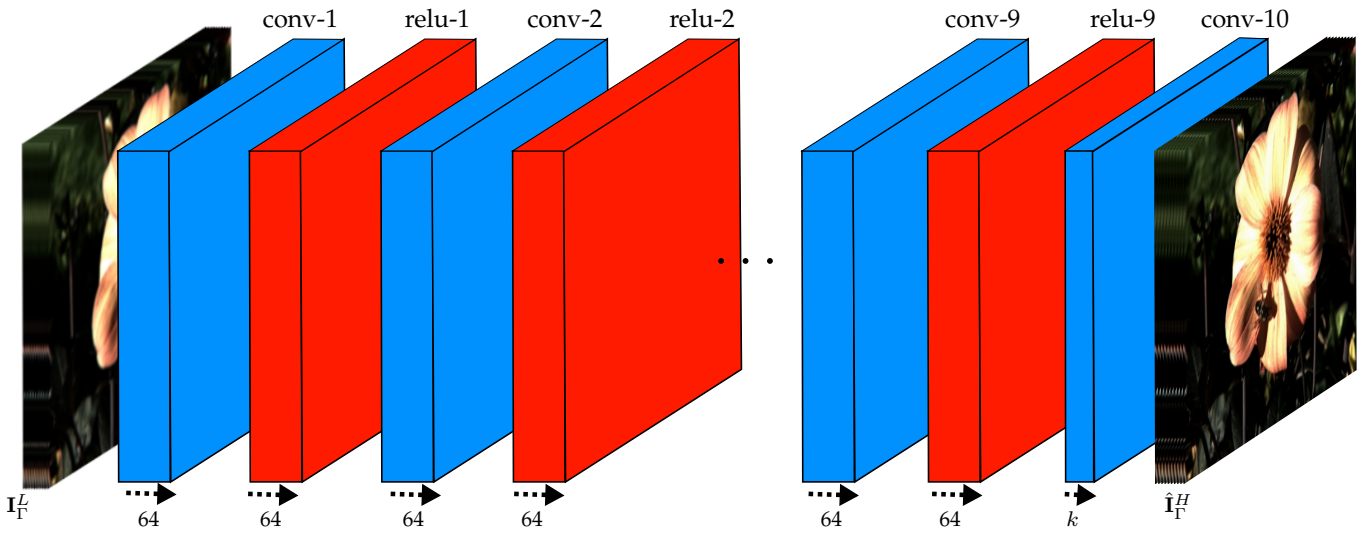


Fig. 4: The proposed network structure which receives a low-resolution light field and restores it using the proposed DCNN.

field $\breve{\mathbf{A}}^H = f\left(\breve{\mathbf{A}}^L\right)$. We then estimate the super-resolved linear dependent part of the light field using

$$\hat{\mathbf{A}}^H = \breve{\mathbf{A}}^H \mathbf{W} \tag{8}$$

The super-resolved low-rank representation of the aligned light field is then derived by the union of the two matrices $\hat{\mathbf{A}}^H$ and $\breve{\mathbf{A}}^H$ *i.e.* $\mathbf{A}^H = \hat{\mathbf{A}}^H \cup \breve{\mathbf{A}}^H$. The super-resolved light field is then reconstructed using

$$\hat{\mathbf{I}}_\Gamma^H = \mathbf{A}^H + \mathbf{E}^L \tag{9}$$

In order to motivate the use of a low-rank prior for the aligned light field we have conducted an experiment where we trained the same CNN architecture depicted in Figure 4 using different ranks. The only modification to this architecture was the number of input channels which is equivalent to the rank k of the aligned light field. For this experiment we have used the same experimental setup described in Section 5.1 and we compute the average PSNR at each rank. Each network was initiated using random weights obtained from a zero mean Gaussian distribution. In all experiments we considered a total of 200 epochs which was found to be sufficient for all networks to converge. It can be seen from Figure 5 that when the rank is too small, the low-rank model will suppress important information which is required to restore the light field. On the other hand, the performance drops when the rank is larger than 20. Increasing the rank k also increases the dimensionality of the problem which makes the CNN more susceptible to over-fitting. The best performance was obtained when training the networks with $k \in [10, 20]$. We have also trained a CNN to restore the low-resolution light field directly (*i.e.* we disabled both alignment and low-rank prior) which achieved an average PSNR of 26.12 dB, which is around 2.5 dB less than the performance achieved by our proposed method with $k = 10$. This result further demonstrates the importance of aligning the light field which reduces the complexity of the function to be learned and therefore manages to achieve better performance.

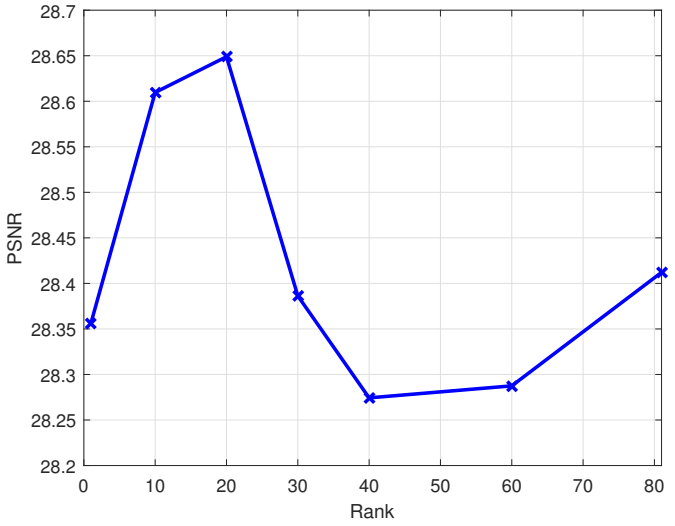
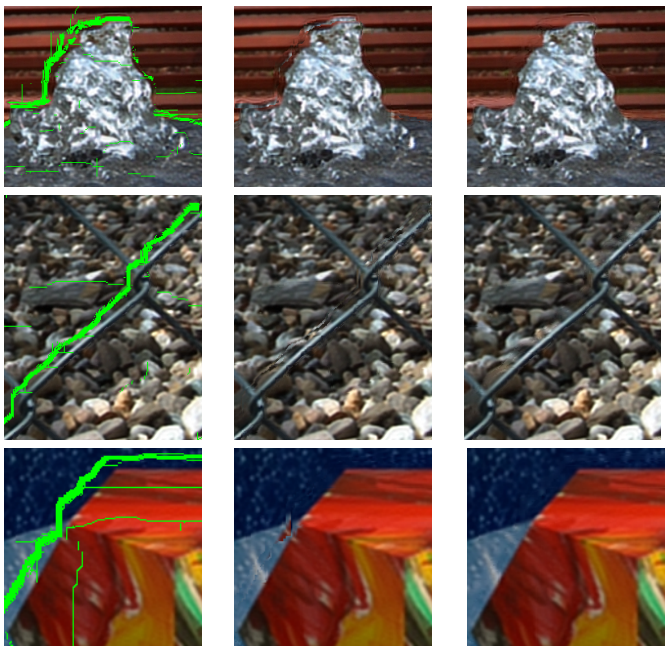


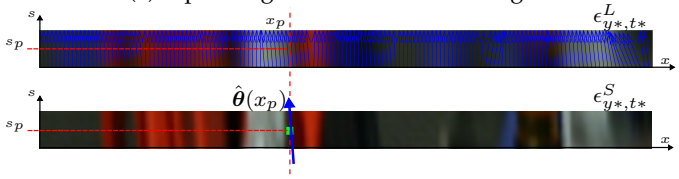
Fig. 5: Analysing the ability of the proposed method to restore the aligned light field when considering different rank values k.

4.3 Light Field Reconstruction

The restored aligned light field $\hat{\mathbf{I}}_\Gamma^H$ has all angular views aligned with the centre view. A naïve approach to recover the original disparities of the restored angular views is to use forward warping. However, as can be seen in the first column of Figure 6a, forward warping is not able to restore all pixels and results in a number of cracks or holes corresponding to occlusions (marked in green). One can use either bicubic interpolation or inverse warping to fill the holes. However, in case of occlusions, the neighbouring pixels may not be well correlated with the missing information, which often results in inaccurate estimations (see Figure 6a second column). More advanced inpainting algorithms [33], [34] can be used to restore each hole separately. However, these methods do not exploit the light field structure and therefore provide inconsistent reconstruction of the same spatial region at different angular views.



(a) Inpainting the cracks marked in green



(b) Diffusion based inpainting

Fig. 6: Filling the missing information caused by occlusion.

In this work, we use a diffusion based inpainting algorithm that estimates the missing pixels by diffusing information available in other views. Similar to the work in [23], we exploit the EPI structure to diffuse information along the dominant orientation of the EPI. However, instead of predicting the orientation of unknown pixels from their spatial neighborhood as done in [23], we exploit the similarity between the low- and super-resolved EPIs and use the structure tensor computed on the low-resolution EPI to guide the inpainting process in the high-resolution EPI.

Without loss of generality we consider the EPI where the dimensions $y*$ and $t*$ are fixed. The case of vertical slices is analogous. We first compute the structure tensor of the low-resolution EPI $\boldsymbol{\epsilon}^L_{y*,t*}$ at coordinates (x, s) using

$$\mathbf{T}(x,s) = \nabla\boldsymbol{\epsilon}^L_{y*,t*}(x,s)\,\nabla\boldsymbol{\epsilon}^L_{y*,t*}(x,s)^\intercal \tag{10}$$

where ∇ stands for the single order gradient computed using the sobel kernel. The authors in [23] compute an average weighting of the columns of $\mathbf{T}(x,s)$ to derive the dominant orientation, where the weights are given by an anisotropy measure. Nevertheless, the anisotropy may fail in regions that are smooth and therefore the weighted average may fail in these regions to estimate the dominant direction of the EPI. This problem becomes an issue when computing these orientations on low-resolution versions of the light fields. Instead, we estimate the orientation at every pixel in the EPI by computing the eigen decomposition of $\mathbf{T}(x,s)$ and choose the direction $\boldsymbol{\theta}(x,s)$ which corresponds to the eigen-vector with the smallest eigen-value. Moreover, driven by the observation that the disparities in a light field are typically small, and considering that the local slope in the EPI is proportional to the disparity, it is reasonable to assume that slopes which correspond to large disparities are less probable to occur. Therefore, to ensure that the tensor driven diffusion is performed along a single coherent direction per column of the EPI and reduce noise, the dominant orientation $\boldsymbol{\theta}(x)$ is computed using the column-wise median of $\boldsymbol{\theta}(x,s)$ whose orientation is in the range $[\frac{\pi}{2}-\alpha, \frac{\pi}{2}+\alpha]$ radians. In all our experiments we set $\alpha = \pi/4\,\mathrm{rad}$. While the dominant orientation vectors are less noisy, we further reduce the noise by applying the Total Variation (TV-L1) denoising [35] on the orientation field $\boldsymbol{\theta}(x)$, which searches to optimize

$$\hat{\boldsymbol{\theta}} = arg\,min_{\hat{\boldsymbol{\theta}}}||\nabla\hat{\boldsymbol{\theta}}||_1 + \lambda||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 \tag{11}$$

where λ is the total variation regularization parameter and was set to 0.5 in our experiments. Figure 6b (top) shows the EPI of the low-resolution light field $\boldsymbol{\epsilon}^L_{y*,t*}$ and the dominant orientations $\hat{\boldsymbol{\theta}}(x)$ marked by blue arrows.

The restored EPI $\boldsymbol{\epsilon}^H_{y*,t*} := \hat{\mathbf{I}}^H(x, y*, s, t*)$ (see Figure 6b (bottom)) has a number of missing pixels (marked in green). Consider that the hole we want to inpaint has coordinates (x_p, s_p). The aim of the proposed diffusion based inpainting algorithm is to propagate known pixels in the orientation $\hat{\boldsymbol{\theta}}(x_p)$ to fill the missing pixels. The diffusion over the EPI $\boldsymbol{\epsilon}^H_{y*,t*}$ evolves as

$$\frac{\partial\boldsymbol{\epsilon}^H_{y*,t*}}{\partial s} = \mathrm{Tr}\left(\hat{\boldsymbol{\theta}}(x_p)\,\hat{\boldsymbol{\theta}}(x_p)^\intercal\,\mathbf{H}(x_p,s_p)\right) \tag{12}$$

where $\mathrm{Tr}(\cdot)$ stands for the trace operator and $\mathbf{H}(x,s)$ denotes the Hessian of $\boldsymbol{\epsilon}^H_{y*,t*}$ at coordinates (x,s). The term $\hat{\boldsymbol{\theta}}(x_p)\,\hat{\boldsymbol{\theta}}(x_p)^\intercal$ is used to enforce the diffusion to occur only in the direction of the isophote eigenvector. The missing pixels are restored iteratively by finding the solution to (12) which is closest to zero. Figure 6a (third column) shows the results attained using the proposed inpainting strategy.

In order to restore all the cracks in the light field we first fix $t*$ to that of the center view and iteratively restore all the horizontal EPIs for all $y* \in [1, Y]$ by solving Eq. (12). This corresponds to filling the cracks for the centre row of the matrix of angular views. We then fix $s*$ to that of the centre view and iteratively restore all the vertical EPIs for all $x* \in [1, X]$ which effectively restores all cracks for the centre column of the matrix of angular views. The remaining propagations are performed row-by-row where each time we restore all pixels within $t* \in [1, Q]$.

4.4 Iterative Back-Projection

One problem with the method proposed in this paper is that after we restore the aligned light field $\hat{\mathbf{I}}^H_\Gamma$ we have to compute inverse warping to restore the original disparities. However, the inverse warping is not able to recover occluded regions and some pixels are displaced by ±1-pixel due to rounding errors. While the former problem is solved using the method described in Section 4.3, the second problem was not yet addressed. Nevertheless, the results illustrated in Tables 1, 2 and Figure 7 indicate that significant performance gains can be achieved even if we do not explicitly cater for distortions caused by rounding errors in the inverse warping process. Nevertheless, these distortions can be corrected using the classical method of iterative back projection [36], which is adopted by several single image super-resolution methods (see [37]) to ensure that the down sampled version of the super-resolved light field is consistent with the observed low-resolution light field. The IBP algorithm iteratively refines the estimated high-resolution light field $\bar{\mathbf{I}}^H_\kappa$ at iteration κ by first back-projecting it into an estimated low-resolution light field $\bar{\mathbf{I}}^L_\kappa$ using

$$\bar{\mathbf{I}}^L_\kappa = \uparrow_\alpha\left(\downarrow_\alpha \mathbf{B}\bar{\mathbf{I}}^H_\kappa\right) \tag{13}$$

where $\downarrow_\alpha$ is a downsampling operator, $\uparrow_\alpha$ is the bicubic upscaling operation, α is the magnification factor and $\mathbf{B}$ is the blurring kernel. The deviation between the LR views found by back-projection and the original LR views is then used to further correct each HR estimated view of the light field as

$$\bar{\mathbf{I}}^H_{\kappa+1} = \bar{\mathbf{I}}^H_\kappa + \left(\mathbf{I}^L - \bar{\mathbf{I}}^L_\kappa\right) \tag{14}$$

The IBP algorithm is initiated by setting $\bar{\mathbf{I}}^H_0 = \hat{\mathbf{I}}^H$ and the iterative procedure terminates when $\kappa = K$. It was observed that significant improvements were achieved in the first few iterations and we therefore set $K = 10$ in our experiments. The restored light field following iterative back-projection is therefore set to $\bar{\mathbf{I}}^H = \bar{\mathbf{I}}^H_K$.

5 Experimental Results

5.1 Evaluation Methodology

The experiments conducted in this paper use both synthetic and real-world light fields from publicly available datasets. We use 98 light fields from the EPFL [38], INRIA[4] and HCI[5] for training. We conducted the tests using light fields from the INRIA and Stanford[6] datasets. We use the Stanford dataset in this evaluation since it has disparities significantly

4. INRIA dataset: https://goo.gl/st8xRt
5. HCI dataset: http://hci-lightfield.iwr.uni-heidelberg.de/
6. Stanford dataset: http://lightfield.stanford.edu/

larger than both INRIA and EPFL light fields, which were captured using plenoptic cameras. Moreover, unlike the HCI dataset, the Stanford light fields capture real world objects. We therefore use this dataset to assess the generalization ability of the algorithms considered in this experiment to light fields which are captured using camera sensors which differ from the ones used for training. While the angular views of the EPFL, HCI and Stanford datasets are available, the light fields in the INRIA dataset were decoded using the method in [39] as mentioned on their website. In all our experiments we consider a 9×9 array of angular views. For computational purposes, the high-resolution images of the Stanford dataset were down-scaled such that the lowest dimension is set to 400 pixels. The high-resolution images of the other datasets were kept unchanged, *i.e.* 512×512 for the HCI light fields and 625×434 for both EPFL and INRIA light fields. Unless otherwise specified, the low-resolution light fields were generated by blurring each high-resolution angular view with a Gaussian filter using a window size of 7 and standard deviation of 1.6, down-sampled to the desired resolution and up-scaled back to the target resolution using bi-cubic interpolation. Unless otherwise specified, the iterative back-projection refinement strategy was disabled to permit a fair comparison to the other state of the art super-resolution methods considered. In all our experiments we set the rank of the aligned light field k and therefore the number of linearly independent components to 10.

We compare the performance of our system against the best performing methods found in our recent work [5], namely the CNN based light field super-resolution algorithm (LF-SRCNN) [11] and both linear subspace projection based methods, PCA+RR and BM+PCA+RR [5]. These methods were retrained using samples from the 98 training light fields mentioned above using training procedures explained in their respective papers. Training the CNN for our method takes roughly one day. Moreover, given that the very deep super-resolution (VDSR) method [31] achieved state-of-the-art performance on single image super-resolution, we apply this method to restore every angular view independently. It is important to mention here that in our previous work [5] we found that BM+PCA+RR significantly outperforms several other light field and single-image super-resolution algorithms including [8], [10], [30], [30], [40], [41], [42]. Due to space constraints we did not provide comparisons against the latter approaches.

5.2 Comparison with existing methods

The results in table 1 and table 2 compare these super-resolution methods in terms of PSNR for magnification factors of $\times 2$ and $\times 3$ respectively. The VDSR algorithm [31] achieves on average a PSNR gain of 0.3 dB and 0.35 dB over bicubic interpolation at magnification factors of $\times 2$ and $\times 3$ respectively. One major limitation of VDSR is that it does not exploit the light field structure where each angular view is being restored independently. The PCA+RR algorithm [5] manages to restore more texture detail and is particularly effective to restore light fields with small disparities, which is the case of the INRIA light fields, but is less effective in case of large disparities. This can be attributed to the fact that PCA+RR does not compensate for disparities and therefore is not able to generalize to light fields containing disparities which were not considered during training, which is the case for the Stanford light fields. The BM+PCA+RR method [5] extends this method by aligning the patch-volumes using block-matching and significantly outperforms PCA+RR, with around 1.2 dB gains in PSNR at both magnification factors, in case of large disparities. The LF-SRCNN method [11], which uses deep learning to restore each light field view independently, was found to achieve a marginal gain over BM+PCA+RR for both magnification factors considered. Nevertheless, our method achieves the best performance with an overall gain of 0.23 dB over the second-best performing algorithm LF-SRCNN.

The results in Figure 7 show the centre views of light fields restored using different light field super-resolution methods when considering a magnification factor of $\times 3$. It can be seen that PCA+RR manages to restore good quality light fields captured by a plenoptic cameras, that have low disparities. Due to the fact that the method does not align the views, it fails when considering light fields with larger disparities (see Lego Knight and Lego Gantry light fields in particular. The performance of BM+PCA+RR improves the generalization and reduces the artifacts attained when restoring light fields with larger disparities. Nevertheless, it evidently fails in restoring the Lego Gantry light field. These subjective results show that LF-SRCNN and our proposed method achieve the best performance, with our method providing sharper light fields (see the bee in the first row, duck's head and feathers in the second row, the helmet of the lego knight in the fourth row and the edges on the camera in the fifth row of Figure 7).

As mentioned in section 4.4, one problem with the proposed method is that the inverse warping is unable to perfectly restore the original disparities of the light field. Nevertheless, the results in tables 1, 2 and Figure 7 clearly show that our proposed method outperforms the other schemes even without the use of the iterative back-projection refinement strategy.

In order to fairly assess the contribution of iterative back-projection, we apply it as a post process for the two best performing methods, namely LF-SRCNN and our proposed scheme LR-LFSR. Tables 3 and 4 show the performance of the two best performing methods with and without iterative back projection as a post-processing step at magnification factors of $\times 2$ and $\times 3$ respectively. It is important to notice that our proposed method outperforms LF-SRCNN when none of them adopts IBP as a post process in terms of both PSNR and SSIM. It can be seen from these results that IBP significantly improves the performance of both methods. Nevertheless, our method followed by iterative back projection as a post process (that is referred to as LR-LFSR-IBP) outperforms all the other methods in terms of both PSNR and SSIM. It achieves PSNR gains of 0.41 dB and 0.31 dB at magnification factors of $\times 2$ and $\times 3$ respectively over LF-SRCNN followed by iterative back projection. It is important to notice that while the performance gain of LR-LFSR over LF-SRCNN without IBP is around 0.23 dB and 0.12 dB at magnification factors of $\times 2$ and $\times 3$ respectively, this gain roughly doubles when both use IBP as a post pro-

TABLE 1: PSNR using different light field super-resolution algorithms when considering a magnification factor of $\times 2$. For clarity bold blue marks the highest and bold red indicates the second highest score.

Light Field Name	Bicubic	PCA+RR [5]	BM+PCA+RR [5]	LF-SRCNN [11]	VDSR [31]	Proposed
Bee 2 (INRIA)	30.1673	**34.1579**	33.4655	**33.8268**	30.4027	33.4915
Dist. Church (INRIA)	24.3059	**26.5071**	26.4571	25.9930	24.5419	**26.7502**
Duck (INRIA)	23.5394	**26.7401**	26.2528	26.0713	23.8371	**26.3777**
Framed (INRIA)	27.5974	**30.7093**	30.4965	30.0697	27.8725	**30.5365**
Fruits (INRIA)	28.4907	31.7884	**32.0002**	31.6820	28.7827	**32.0789**
Mini (INRIA)	27.6332	**30.4751**	30.0867	29.8941	27.9175	**30.1601**
Rose (INRIA)	33.5436	**37.0791**	**36.9566**	36.8245	33.7943	36.8416
Amethyst (STANFORD)	30.5227	**32.5139**	**32.4262**	32.2953	30.8360	32.2737
Bracelet (STANFORD)	26.4662	23.8183	28.2356	**28.8858**	26.8523	**29.4046**
Chess (STANFORD)	30.2895	31.9292	**32.5708**	32.1922	30.6313	**32.6123**
Eucalyptus (STANFORD)	30.7865	32.4162	**32.4900**	32.1989	31.0431	**32.6205**
Lego Gantry (STANFORD)	27.6235	28.0729	28.7230	**29.8086**	27.9998	**29.8112**
Lego Knights (STANFORD)	27.3794	27.8457	**29.4664**	**29.5354**	27.7446	29.3177

TABLE 2: PSNR using different light field super-resolution algorithms when considering a magnification factor of $\times 3$. For clarity bold blue marks the highest and bold red indicates the second highest score.

Light Field Name	Bicubic	PCA+RR [5]	BM+PCA+RR [5]	LF-SRCNN [11]	VDSR [31]	Proposed
Bee 2 (INRIA)	27.8623	31.2436	31.1886	**31.3945**	28.2457	**31.2545**
Dist. Church (INRIA)	23.3138	**24.7103**	24.6220	24.5874	23.7707	**24.6535**
Duck (INRIA)	22.0702	23.9844	23.8083	**24.0623**	22.5350	**24.1549**
Framed (INRIA)	26.1627	28.0771	**28.1587**	27.9157	26.8462	**28.2954**
Fruits (INRIA)	26.5269	29.0908	29.1969	**29.2100**	26.9021	**29.5297**
Mini (INRIA)	26.3035	**28.4265**	28.3212	28.1731	26.7192	**28.4009**
Rose (INRIA)	31.7687	**34.4692**	**34.5754**	34.3064	32.0424	34.3392
Amethyst (STANFORD)	29.0665	**31.0848**	**31.1184**	30.5971	29.5618	30.4628
Bracelet (STANFORD)	24.1221	22.2654	25.3946	**26.1013**	24.3484	**26.2712**
Chess (STANFORD)	27.3679	29.5207	29.6773	**30.0278**	27.6514	**30.1485**
Eucalyptus (STANFORD)	29.2136	**31.3459**	**31.3547**	30.9433	29.4650	31.1772
Lego Gantry (STANFORD)	24.6721	25.3941	25.6383	**26.8054**	24.7923	**26.9466**
Lego Knights (STANFORD)	24.0182	24.4944	25.5768	**26.0771**	24.1256	**26.1358**

cessing step. This indicates that since LF-SRCNN processes each view independently, IBP only corrects inconsistencies between the low-resolution and restored light fields. Apart from this distortion, LR-LFSR-IBP corrects the distortions caused by the inverse warping process which provides light fields which are more visually pleasing and with smoother transitions across views. Supplementary multimedia files uploaded on ScholarOne show the pseudo videos of a number of restored light fields and show that the restored light fields are sharper, angularly coherent and that the proposed method is able to restore real-world plenoptic light fields. Supplementary material is available on the project's website[7] while the code of the LR-LFSR will be made available upon publication.

5.3 Analysis of non-Lambertian Surfaces

The variations across the angular views, and therefore the rank of the light field, is not only affected by disparities and occlusions as stated above, but is also influenced by specularity, reflection and refraction from curved surfaces, transparency, subsurface scattering and other non-Lambertian light phenomena. Figure 8 shows the performance of the proposed method when restoring light fields containing such lighting phenomenas. It can be seen that the light fields restored using our proposed method (denoted by $\bar{\mathbf{I}}^H$) are of significant higher quality compared to the low-resolution light field $\mathbf{I}^L$, even in regions that are considered as non-Lambertian. Moreover, the EPIs show that the restored light field is angularly coherently and preserves its geometrical structure even in presence of non-Lambertian lighting phenomena.

7. https://goo.gl/8DDsDi

5.4 Complexity Analysis

The complexity of the proposed method is mainly affected by the computation of the optical flows used to align the light field (SIFT Flow in our case), SVD decomposition that is used to estimate a low-rank model of the aligned light field, identifying the linearly-independent components of the low-rank model, the restoration of the linearly independent components of the light field and the inpainting of the missing pixels. The Sift Flow used to align all the n views to the center view is reported in [27] to have a time complexity of the order $O(nm\log(\sqrt{m}))$, where m represents the number of pixels in each view. The SVD decomposition has a time complexity of the order $O(n^2m)$ while the complexity of the QR decomposition used to find the linear independent components is of the order $O(n^3)$ where $n \ll m$. The feed-forward CNN used to restore the aligned linearly independent components has a fixed depth and with and its complexity is mainly dependent on the target spatial resolution of the light field. This implies that the restoration process has a time complexity of the order $O(m)$ Finally, the inpainting process, which is applied to restore the missing pixels, is only applied on a fraction of

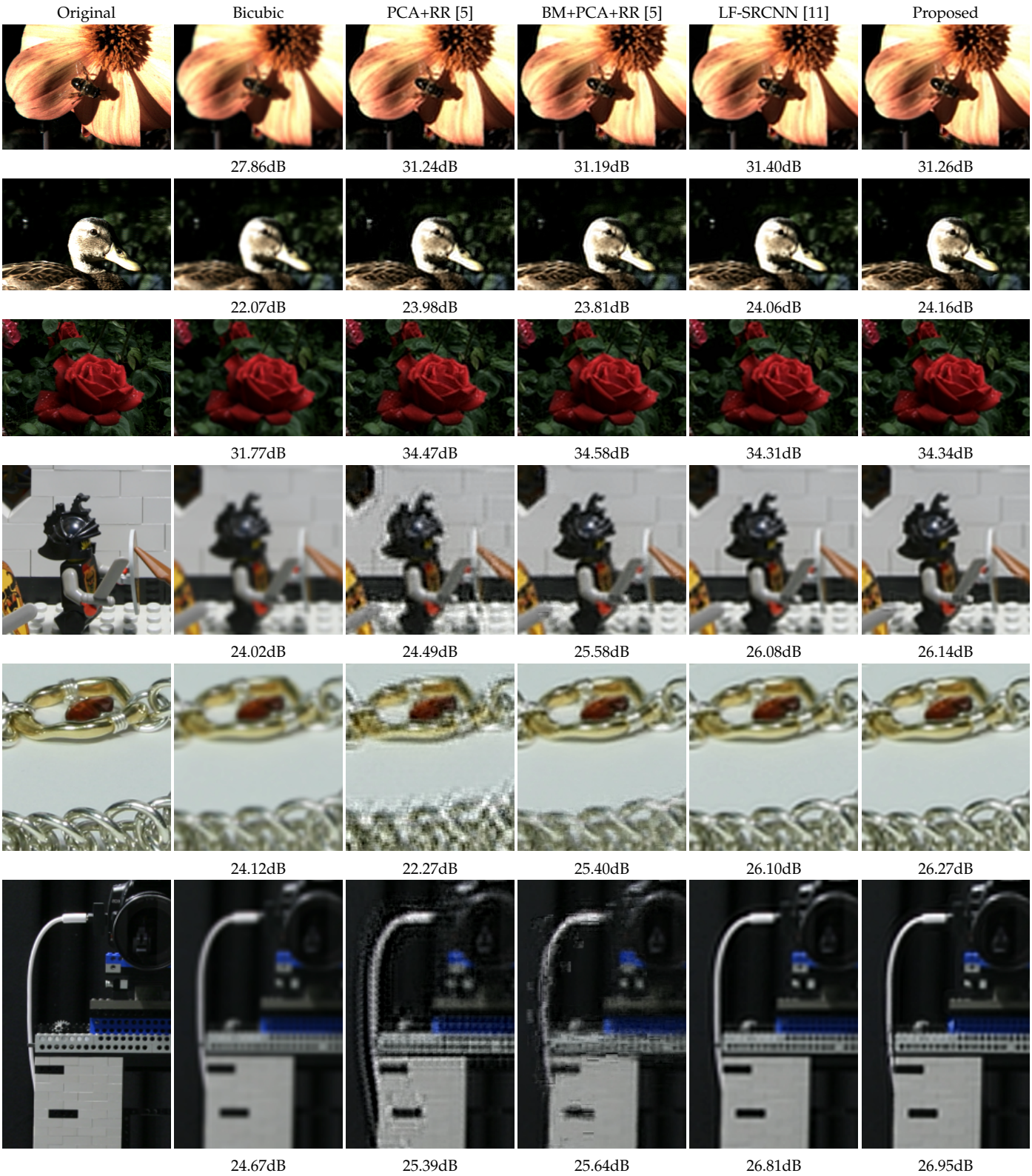


Fig. 7: Restored center view using different light field super-resolution algorithms. These are best viewed in color and by zooming on the views. Underneath each image we show the PSNR values

.

the number of pixels contained in a light field. We therefore estimate the complexity of the inpainting process to be of the order $O(\gamma mn)$, where $\gamma \leq 1$ represents the probability of missing pixels which is typically in the range between 0.01 and 0.05.

A qualitative assessment of the complexity of different light field super-resolution methods considered in this work is summarized in Table 5. These methods were implemented

TABLE 3: PSNR (SSIM in parenthesis) quality measures obtained with the best two performing methods at a magnification of $\times 2$ with and without iterative back projection as a post process. For clarity bold blue marks the highest and bold red indicates the second highest score.

Light Field Name	Bicubic	LF-SRCNN	LF-SRCNN-IBP	LR-LFSR	LR-LFSR-IBP
Bee 2 (INRIA)	30.1673 (0.8975)	33.8268 (0.9380)	**34.6703** (**0.9397**)	33.4915 (0.9329)	**34.9938** (**0.9388**)
Dist. Church (INRIA)	24.3059 (0.7173)	25.9930 (0.7950)	26.4473 (0.8059)	**26.7502** (**0.8198**)	**27.0778** (**0.8247**)
Duck (INRIA)	23.5394 (0.8104)	26.0713 (0.8938)	**26.9118** (**0.9042**)	26.3777 (0.8988)	**27.5095** (**0.9148**)
Framed (INRIA)	27.5974 (0.8690)	31.0592 (0.9135)	**31.5897** (**0.9168**)	30.5365 (0.9079)	**31.8182** (**0.9210**)
Fruits (INRIA)	28.4907 (0.8478)	31.6820 (0.9157)	**32.5159** (0.9230)	32.0789 (**0.9252**)	**33.2449** (**0.9332**)
Mini (INRIA)	27.6332 (0.7666)	29.8941 (0.8365)	**30.3944** (0.8455)	30.1601 (**0.8506**)	**30.8949** (**0.8592**)
Rose (INRIA)	33.5436 (0.8859)	36.8245 (0.9387)	**37.4991** (**0.9420**)	36.8416 (**0.9420**)	**37.8970** (**0.9461**)
Amethyst (STANFORD)	30.5227 (0.8959)	32.2953 (0.9287)	**33.3249** (0.9342)	32.2737 (**0.9349**)	**33.7299** (**0.9397**)
Bracelet (STANFORD)	26.4662 (0.9108)	28.8858 (0.9108)	**30.1670** (**0.9253**)	29.4046 (0.9251)	**30.7842** (**0.9331**)
Chess (STANFORD)	30.2895 (0.9104)	32.1922 (0.9402)	**33.3866** (0.9464)	32.6123 (**0.9486**)	**34.0083** (**0.9519**)
Eucalyptus (STANFORD)	30.7865 (0.8998)	32.1989 (0.9218)	**32.8917** (0.9258)	32.6205 (**0.9316**)	**33.3725** (**0.9332**)
Lego Gantry (STANFORD)	27.6235 (0.8718)	29.8086 (0.9088)	**30.9896** (**0.9177**)	29.8112 (0.9156)	**31.4715** (**0.9242**)
Lego Knights (STANFORD)	27.3794 (0.8463)	29.5354 (0.8977)	**30.9744** (**0.9115**)	29.3177 (0.8991)	**31.1533** (**0.9125**)

TABLE 4: PSNR (SSIM in parenthesis) quality measures obtained with the best two performing methods at a magnification of $\times 3$ with and without iterative back projection as a post process. For clarity bold blue marks the highest and bold red indicates the second highest score.

Light Field Name	Bicubic	LF-SRCNN	LF-SRCNN-IBP	LR-LFSR	LR-LFSR-IBP
Bee 2 (INRIA)	27.8623 (0.8607)	31.3945 (0.9107)	**32.4628** (**0.9129**)	31.2545 (0.9088)	**32.7379** (**0.9140**)
Dist. Church (INRIA)	23.3138 (0.6724)	24.5874 (0.7367)	**25.1374** (**0.7481**)	24.6535 (0.7441)	**25.1214** (**0.7508**)
Duck (INRIA)	22.0702 (0.7503)	24.0623 (0.8357)	**24.7215** (**0.8467**)	24.1549 (0.8397)	**25.0135** (**0.8564**)
Framed (INRIA)	26.1627 (0.8356)	27.9157 (0.8709)	**28.8273** (0.8720)	28.2954 (**0.8788**)	**29.4067** (**0.8793**)
Fruits (INRIA)	26.5269 (0.7990)	29.2100 (0.8656)	**30.1651** (0.8709)	29.5438 (**0.8783**)	**30.7763** (**0.8852**)
Mini (INRIA)	26.3035 (0.7211)	28.1731 (0.7823)	**28.6895** (0.7892)	28.4009 (**0.7956**)	**29.0910** (**0.8042**)
Rose (INRIA)	31.7687 (0.8457)	34.3064 (0.9012)	**34.9356** (**0.9070**)	34.3392 (0.9053)	**35.4656** (**0.9125**)
Amethyst (STANFORD)	29.0665 (0.8676)	30.5971 (0.9003)	**31.5229** (**0.9054**)	30.4628 (0.9047)	**31.7441** (**0.9091**)
Bracelet (STANFORD)	24.1221 (0.7710)	26.1013 (0.8452)	**26.9047** (0.8574)	26.2712 (**0.8584**)	**27.2311** (**0.8708**)
Chess (STANFORD)	27.3679 (0.8574)	30.0279 (0.9045)	**31.0846** (0.9067)	30.1485 (**0.9122**)	**31.3733** (**0.9134**)
Eucalyptus (STANFORD)	29.2136 (0.8751)	30.9433 (0.9016)	**31.5263** (0.9017)	31.1772 (**0.9098**)	**31.8815** (**0.9093**)
Lego Gantry (STANFORD)	24.6721 (0.8061)	26.8054 (0.8519)	**27.5996** (0.8560)	26.9466 (**0.8614**)	**27.8655** (**0.8665**)
Lego Knights (STANFORD)	24.0182 (0.7512)	26.0771 (0.8238)	**27.6550** (**0.8293**)	26.1358 (0.8283)	**27.7168** (**0.8363**)

using MATLAB with code provided by the authors and tested on the same computer with Interl Core (TM)i7, 64-bit Windows 10 operating system, 32-GByte of RAM and a Tital GTX1080Ti GPU. It can be seen that our method ranks second in terms of complexity with LF-SRCNN achieving the best performance. We must however mention that the optical flow process used in our method dominates the complexity of our proposed method.

TABLE 5: Processing time of different light field super-resolution algorithms at different magnification factors.

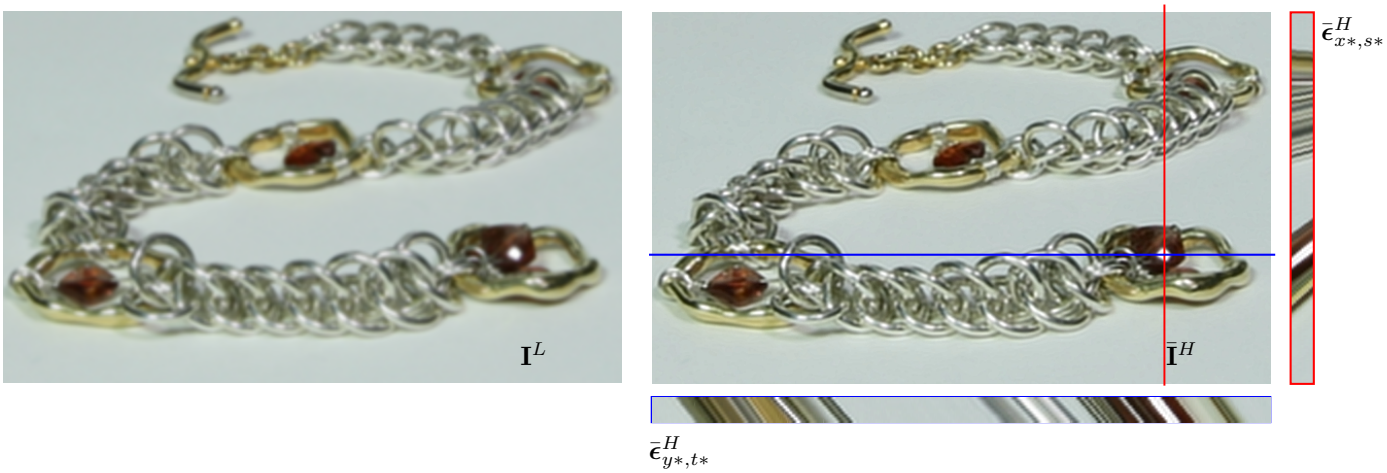
Algorithm	$\times 2$	$\times 3$	$\times 4$
PCARR	3 min.	3 min.	3 min.
BM-PCARR	22 min.	23 min.	23 min.
LF-SRCNN	33 sec.	33 sec.	33 sec.
Proposed	12 min.	12 min.	12 min.

5.5 Digital Refocusing

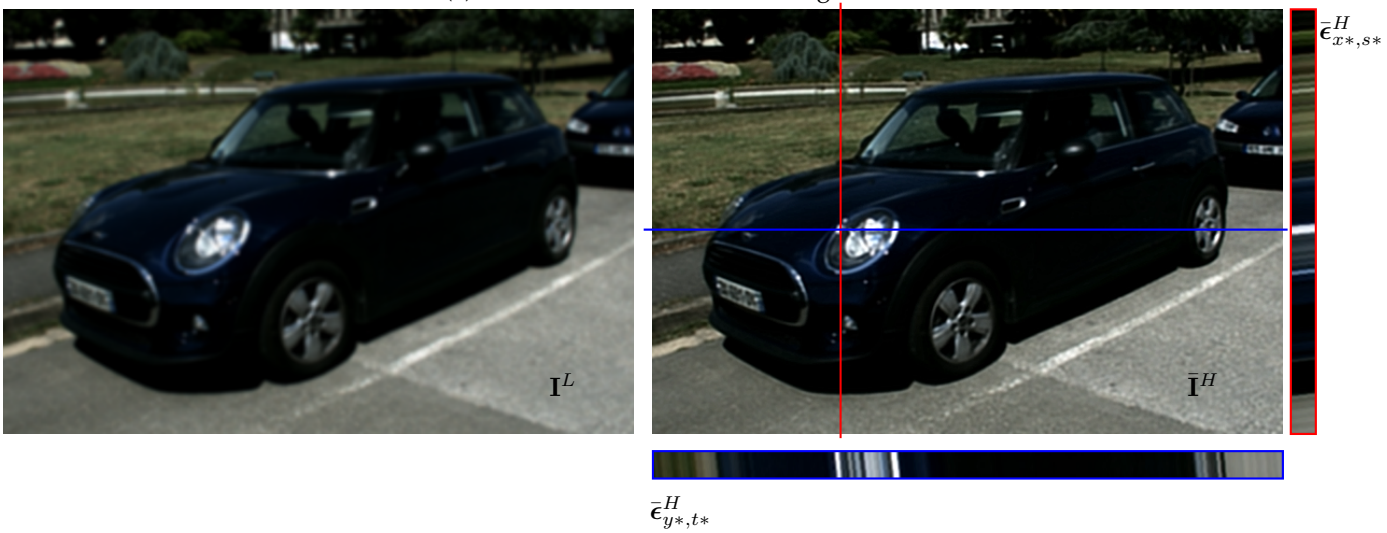
The geometric structure of the light field can be exploited to allow to digitally refocus at post production. This is one of the most important features provided by light fields that is not possible by conventional cameras and is directly impacted by the quality of the light field. The results in figure 9 show a number of refocused images obtained from light fields restored using the LF-SRCNN [11] and our proposed method, where the images are refocused using the Light Field Toolbox [43]. These results show that refocused images generated by light fields restored using our method are superior to those restored using LF-SRCNN (See the sharper eyes of the Duck in the first row and the detail on the helmet of the Lego soldier in Figure 9). Moreover, it can be seen that images refocused using light fields restored using our method are sharper, even when consider non-Lambertian surfaces.

6 Conclusion

In this paper, we have proposed a novel spatial light field super-resolution algorithm able to reconstruct high quality coherent light fields. We have shown that the information in a light field can be efficiently compacted by aligning the angular views using optical flow followed by low-rank matrix approximation. The low rank approximation of the aligned light field gives an embedding in a lower dimensional space which is super-resolved using deep learning. All aligned views of the high-resolution light field can be reconstructed from the super-resolved embedding by simple linear combinations. These views are then inverse warped to restore the disparities of the original light field. Holes corresponding to dis-occlusions or cracks resulting from



(a) EPI of the restored Bracelet Light Field.



(a) EPI of the restored Mini Light Field.

Fig. 8: Analysing the EPI geometry of Light Fields restored using our proposed method on non-Lambertian surfaces.

the inverse warping are filled in using a novel diffusion based inpainting algorithm which diffuses known pixels in the EPI along dominant orientations computed in the low-resolution EPI.

Extensive simulations show that the proposed method manages to generalize well, i.e. manages to successfully restore light fields whose disparities are considerably different from those used during training. These results also show that our proposed method is competitive and most of the time superior to existing state-of-the-art light field super-resolution algorithms, including a recent approach which adopts deep learning to restore each view independently. One major limitation of the proposed scheme is that the inverse warping process is not able to restore the original disparities and produces some distortion caused by rounding errors. We proposed here to use the classical iterative back-projection as a post processing step. Simulation results clearly show the benefit of using IBP as a post processing of the super-resolved light field and demonstrate that the proposed method with IBP achieves the best performance, outperforming LF-SRCNN followed by IBP by 0.4 dB. Future work will involve in extending this method to perform angular super-resolution.

References

[1] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," Stanford University, Tech. Rep., Apr. 2005. [Online]. Available: http://graphics.stanford.edu/papers/lfcamera/

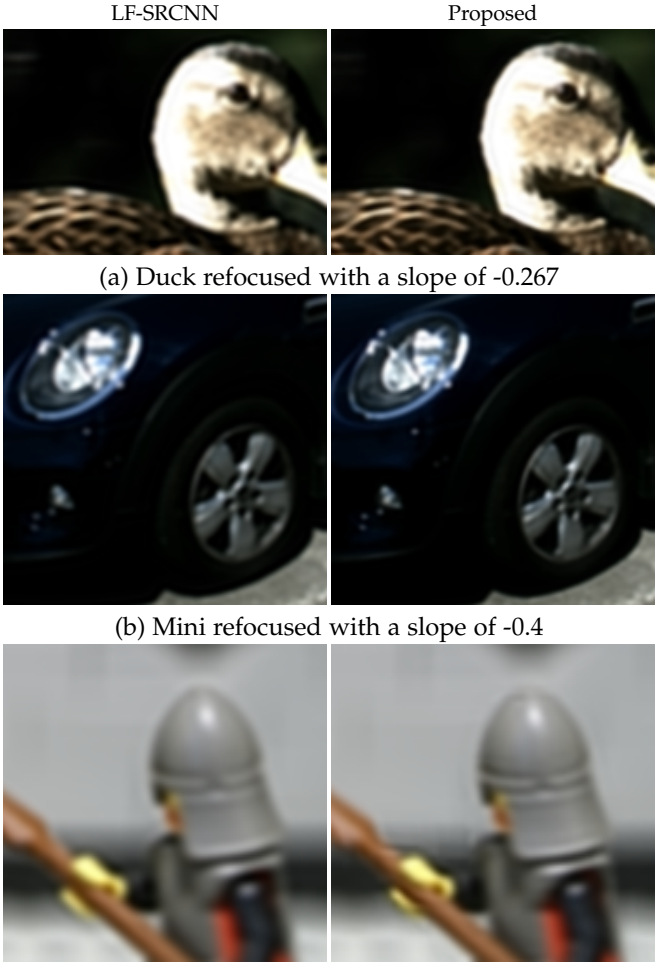
[2] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 765–776, Jul. 2005. [Online]. Available: http://doi.acm.org/10.1145/1073204.1073259

[3] G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field image processing: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[4] C.-K. Liang and R. Ramamoorthi, "A light transport framework for lenslet light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 16:1–16:19, Mar. 2015. [Online]. Available: http://doi.acm.org/10.1145/2665075

[5] R. A. Farrugia, C. Galea, and C. Guillemot, "Super resolution of light field images using linear subspace projection of patch-volumes," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[6] A. Levin, W. Freeman, and F. Durand, "Understanding camera trade-offs through a bayesian analysis of light field projections," in *European Conference on Computer Vision (ECCV)*, Oct. 2008.



(a) Duck refocused with a slope of -0.267

(b) Mini refocused with a slope of -0.4

(c) Lego Knights refocused with a slope of -0.2

Fig. 9: Light fields refocused using different slopes. The light fields were restored using LF-SRCNN method [11] and the method proposed in this paper.

[7] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and superresolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 972–986, May 2012.

[8] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocussing using a gmm light field patch prior," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 22–28.

[9] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, March 2014.

[10] Y. Yoon, H.-G. Jeon, D. Yoo, J.-Y. Lee, and I. S. Kweon, "Learning a deep convolutional network for light-field image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 57–65.

[11] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon, "Light-field image super-resolution using convolutional neural network," *IEEE Signal Processing Letters*, vol. 24, no. 6, pp. 848–852, June 2017.

[12] M. S. K. Gul and B. K. Gunturk, "Spatial and angular resolution enhancement of light fields using convolutional neural networks," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2146–2159, May 2018.

[13] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2016)*, vol. 35, no. 6, 2016.

[14] G. Wu, M. Zhao, L. Wang, Q. Dai, T. Chai, and Y. Liu, "Light field reconstruction using deep convolutional network on epi," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[15] T.-C. Wang, J.-Y. Zhu, N. K. Kalantari, A. A. Efros, and R. Ramamoorthi, "Light field video capture using a learning-based hybrid imaging system," *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, vol. 36, no. 4, 2017.

[16] Y. Wang, Y. Liu, W. Heidrich, and Q. Dai, "The light field attachment: Turning a dslr into a light field camera using a low budget camera ring," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 10, pp. 2357–2364, Oct 2017.

[17] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011. [Online]. Available: http://doi.acm.org/10.1145/1970392.1970395

[18] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 763–770.

[19] X. Jiang, M. L. Pendu, R. A. Farrugia, and C. Guillemot, "Light field compression with homography-based low rank approximation," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[20] S. M. Seitz and K. N. Kutulakos, "Plenoptic image editing," in *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, Jan 1998, pp. 17–24.

[21] H. Ao, Y. Zhang, A. Jarabo, B. Masia, Y. Liu, D. Gutierrez, and Q. Dai, "Light field editing based on reparameterization," in *Advances in Multimedia Information Processing – PCM 2015*, Y.-S. Ho, J. Sang, Y. M. Ro, J. Kim, and F. Wu, Eds. Cham: Springer International Publishing, 2015, pp. 601–610.

[22] F. L. Zhang, J. Wang, E. Shechtman, Z. Y. Zhou, J. X. Shi, and S. M. Hu, "Plenopatch: Patch-based plenoptic image manipulation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 5, pp. 1561–1573, May 2017.

[23] O. Frigo and C. Guillemot, "Epipolar plane diffusion: An efficient approach for light field editing," in *British Machine Vision Conference (BMVC)*, London, France, Sep. 2017.

[24] M. Levoy and P. Hanrahan, "Light field rendering," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[25] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '96. New York, NY, USA: ACM, 1996, pp. 43–54. [Online]. Available: http://doi.acm.org/10.1145/237170.237200

[26] B. K. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, no. 1, pp. 185 – 203, 1981. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0004370281900242

[27] C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 5, pp. 978–994, May 2011.

[28] Y. Hu, R. Song, and Y. Li, "Efficient coarse-to-fine patch match for large displacement optical flow," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 5704–5712.

[29] Y. Li, D. Min, M. S. Brown, M. N. Do, and J. Lu, "Spm-bp: Sped-up patchmatch belief propagation for continuous mrfs," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 4006–4014.

[30] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, Feb 2016.

[31] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1646–1654.

[32] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS-10)*, Y. W. Teh and D. M. Titterington, Eds.,

vol. 9, 2010, pp. 249–256. [Online]. Available: http://www.jmlr.org/proceedings/papers/v9/glorot10a/glorot10a.pdf
[33] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *Trans. Img. Proc.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004. [Online]. Available: http://dx.doi.org/10.1109/TIP.2004.833105
[34] Z. Xu and J. Sun, "Image inpainting by patch propagation using patch sparsity," *IEEE Transactions on Image Processing*, vol. 19, no. 5, pp. 1153–1165, May 2010.
[35] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1, pp. 259 – 268, 1992. [Online]. Available: http://www.sciencedirect.com/science/article/pii/016727899290242F
[36] M. Irani and S. Peleg, "Improving resolution by image registration," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp. 231 – 239, 1991. [Online]. Available: http://www.sciencedirect.com/science/article/pii/104996529190045L
[37] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 349–356.
[38] M. Rerabek and T. Ebrahimi, "New light field image dataset," in *Int. Conf. on Quality of Experience*, 2016.
[39] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration and rectification for lenselet-based plenoptic cameras," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1027–1034.
[40] T. Peleg and M. Elad, "A statistical prediction model based on sparse representations for single image super-resolution," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2569–2582, June 2014.
[41] K. Zhang, B. Wang, W. Zuo, H. Zhang, and L. Zhang, "Joint learning of multiple regressors for single image super-resolution," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 102–106, Jan 2016.
[42] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *2013 IEEE International Conference on Computer Vision*, Dec 2013, pp. 1920–1927.
[43] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Linear volumetric focus for light field cameras," *ACM Trans. Graph.*, vol. 34, no. 2, pp. 15:1–15:20, Mar. 2015.

ACKNOWLEDGMENTS

This project has been supported in part by the EU H2020 Research and Innovation Programme under grant agreement No 694122 (ERC advanced grant CLIM).

Reuben A. Farrugia (S04, M09, SM17) received the first degree in Electrical Engineering from the University of Malta, Malta, in 2004, and the Ph.D. degree from the University of Malta, Malta, in 2009. In January 2008 he was appointed Assistant Lecturer with the same department and is now a Senior Lecturer. He has been in technical and organizational committees of several national and international conferences. In particular, he served as General-Chair on the IEEE Int. Workshop on Biometrics and Forensics (IWBF) and as Technical Programme Co-Chair on the IEEE Visual Communications and Image Processing (VCIP) in 2014. He has been contributing as a reviewer of several journals and conferences, including IEEE Transactions on Image Processing, IEEE Transactions on Circuits and Systems for Video and Technology and IEEE Transactions on Multimedia. On September 2013 he was appointed as National Contact Point of the European Association of Biometrics (EAB).

Christine Guillemot IEEE fellow, is Director of Research at INRIA, head of a research team dealing with image and video modeling, processing, coding and communication. She holds a Ph.D. degree from ENST (Ecole Nationale Superieure des Telecommunications) Paris, and an Habilitation for Research Direction from the University of Rennes. From 1985 to Oct. 1997, she has been with FRANCE TELECOM, where she has been involved in various projects in the area of image and video coding for TV, HDTV and multimedia. From Jan. 1990 to mid 1991, she has worked at Bellcore, NJ, USA, as a visiting scientist. She has (co)-authored 24 patents, 9 book chapters, 60 journal papers and 140 conference papers. She has served as associated editor (AE) for the IEEE Trans. on Image processing (2000-2003), for IEEE Trans. on Circuits and Systems for Video Technology (2004-2006) and for IEEE Trans. On Signal Processing (2007-2009). She is currently AE for the Eurasip journal on image communication, IEEE Trans. on Image Processing (2014-2016) and member of the editorial board for the IEEE Journal on selected topics in signal processing (2013-2015). She is a member of the IEEE IVMSP technical committee.