Super Resolution of Light Field Images using Linear Subspace Projection of Patch-Volumes

Reuben A Farrugia, Christian Galea, Christine Guillemot

Super Resolution of Light Field Images using Linear Subspace Projection of Patch-Volumes

Reuben A. Farrugia, *Member, IEEE,* Christian Galea, *Student Member, IEEE* and Christine Guillemot, *Fellow, IEEE,*

***Abstract*—Light field imaging has emerged as a very promising technology in the field of computational photography. Cameras are becoming commercially available for capturing real-world light fields. However, capturing high spatial resolution light fields remains technologically challenging, and the images rendered from real light fields have today a significantly lower spatial resolution compared to traditional 2D cameras. This paper describes an example-based super-resolution algorithm for light fields, which allows the increase of the spatial resolution of the different views in a consistent manner across all sub-aperture images of the light field. The algorithm learns linear projections between subspaces of reduced dimension in which reside patch-volumes extracted from the light field. The method is extended to cope with angular super-resolution, where 2D patches of intermediate sub-aperture images are approximated from neighbouring sub-aperture images using multivariate ridge regression. Experimental results show significant quality improvement when compared to state-of-the-art single-image super-resolution methods applied on each view separately, as well as when compared to a recent light field super-resolution techniques based on deep learning.**

***Index Terms*—Light-fields, Super-resolution, Learning, Dimensionality reduction.**

I. Introduction

Light fields are densely sampled high-dimensional signals containing information about the light rays interacting with the scene. Compared to classical 2D images, light fields capture the intensity values along each ray and not only the sum of intensities of rays reaching each image point. This rich description of a scene enables advanced creation of images with a number of functionalities such as refocusing, extended depth of field, different view point rendering and depth reconstruction which can be very useful for image/video post-production editing and cinematic virtual reality applications [1], [2]. Many light fields acquisition devices have been recently designed, going from arrays of cameras capturing the scene from slightly different viewpoints [2], to single cameras mounted on moving gantries and plenoptic cameras.

Plenoptic cameras, now commercially available, use an array of micro-lenses placed between the sensor and the main lens, to acquire both spatial and angular information of light rays from a single capture [1]. However, since the sensor resolution is limited, it is difficult to have both a dense angular and spatial light field sampling. 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. This trade-off between angular and spatial resolution leads to a significantly lower spatial resolution compared to traditional 2D cameras [1].

Research effort has already been dedicated to this angular and spatial resolution trade-off in light field imaging. A first category of approaches consists in super-resolving the rendered images, *i.e*. enhancing spatial super-resolution, as in [3] where the authors simultaneously estimate a super-resolved depth map and the all-in-focus image. In the same vein, a light field reconstruction approach is proposed in [4], where the de-multiplexed sub-aperture images are first interpolated with barycentric interpolation to adapt to the hexagonal layout of the micro-lenses, and then refined using pixels of neighboring views using ray interpolation. The rendered image is further enhanced using the dictionary-based super-resolution method proposed in [5] for 2D images.

A second category of methods directly super-resolve the light field to enhance both the spatial and angular resolution. Based on an image formation model of the plenoptic camera, the authors in [6] cast the high-resolution (HR) light field estimation in a Bayesian framework assuming a depth-dependent blurring kernel. A method based on a variational framework is proposed in [7] to synthesize super-resolved novel views with the help of disparity information estimated on the epipolar plane images. A patch-based technique is proposed in [8] to enhance the spatial resolution of light fields. The high resolution 4D-patches are estimated using a linear minimum mean square error (LMMSE) estimator, assuming a disparity-dependent Gaussian Mixture Model (GMM) for the patch structure. More recently, the authors in [9] adopt a deep convolutional neural network (DCNN) to perform both spatial and angular super-resolution. This method first employs a spatial DCNN similar to [10] followed by an angular DCNN which synthesizes intermediate views. A cascade of two DCNNs is used in [11] which exploits disparity information to synthesize intermediate sub-aperture images. A sparse representation approach that can be extended for light-field super-resolution was presented in [12].

In this paper, we propose an example-based spatial super-resolution technique, which, to maintain consistency across all sub-aperture images of the light field, operates on 3D stacks (called *patch-volumes*) of 2D-patches, extracted from the different sub-aperture images. The patches forming the 3D

R.A. Farrugia is with the Department of Communications and Computer Engineering, University of Malta, Malta, e-mail: (reuben.farrugia@um.edu.mt).

C. Galea is with the Department of Communications and Computer Engineering, University of Malta, Malta, e-mail: (christian.galea.09@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).

stack are either co-located patches or best matches across sub-aperture images. A dictionary of *examples* is first constructed by extracting, from a training set of high- and low- resolution light fields, pairs of high- and low-resolution *patch-volumes*. These *patch-volumes* are of very high dimension ($q \times q \times n$), where n is the number of sub-aperture images and $q \times q$ is the size of each 2D-patch. Nevertheless, they contain a lot of redundant information, hence actually lie on subspaces of lower dimension. The low- and high-resolution *patch-volumes* of each pair can therefore be projected on their respective low- and high-resolution subspaces. In the sequel, the projection is performed using Principal Component Analysis (PCA). The dictionary of pairs of projected *patch-volumes* (the *examples*) map locally the relation between the high-resolution *patch-volumes* and their low-resolution (LR) counterparts. A linear mapping function is then learned, using Multivariate Ridge Regression (RR), between the subspaces of the low- and high-resolution *patch-volumes*. Each overlapping *patch-volume* of the low-resolution light field can then be super-resolved by a straight application of the learned mapping function.

The above method, which will be referred to as PCA+RR in the sequel, assumes that the 2D collocated patches extracted from all sub-aperture images to form a given *patch-volume*, are well aligned. This may not be the case when large disparities exist across sub-aperture images, depending on the depth of the scene and of the capturing device. For light fields exhibiting large disparities, the above method is further improved by using block matching (BM) to form the *patch-volumes* with the best-matching patches across all sub-aperture images, instead of simply taking collocated patches. An iterative procedure is proposed where a different sub-aperture image is chosen as anchor at each iteration to form the *patch-volumes*. This method will be referred to as BM+PCA+RR.

By exploiting the light field structure, where sub-aperture images represent different viewpoints from the same scene, it becomes evident that a global interpolation function to cater for all views will be sub-optimal. The DCNN scheme presented in [9] models the angular super-resolution function based on the available neighbouring sub-aperture images independently from the viewpoint being approximated. Instead, we model the angular super-resolution problem using multi-linear models, *i.e.* a linear model for every missing sub-aperture image to be synthesized. This is done by exploiting the local neighbouring sub-aperture image and also the angular location of the sub-aperture image being approximated.

Experimental results with both synthetic light fields (from the HCI dataset) and real light fields (from the Stanford and the INRIA datasets) show that the proposed method outperforms state-of-the-art single-image super-resolution methods [13], [14], [15] applied on each sub-aperture image separately, and significantly outperfoms prior light field super-resolution algorithms for both spatial [8], [9] and angular [9], [11] super-resolution. Average PSNR gains of 3.7dB, 4.2dB and 2.1dB on the HCI, Stanford and INRIA datasets respectively were obtained for spatial super-resolution using BM+PCA+RR compared to the recent DCNN scheme [9] which obtained the second best results in our experiments. Moreover, subjective results clearly show that the proposed angular super-resolution method provides sharper synthesized intermediate sub-aperture images when compared to those obtained using the DCNN based schemes [9], [11].

The rest of the paper is organized as follows. After introducing the notations in Section II, we describe the different steps of the proposed algorithm in Section III. Section IV discusses the simulation results with different types of light fields and provide the final concluding remarks in Section VI.

II. Light Fields Notation

We consider here the simplified 4D representation of light fields called 4D light field in [16] and lumigraph in [17], describing the radiance along rays by a function $L(x, y, s, t)$ where the pairs (x, y) and (s, t) respectively represent spatial and angular coordinates respectively. The light field can be seen as capturing an array of viewpoints (called sub-aperture images) of the scene with varying angular coordinates (s, t). The different views will be denoted here by $\mathbf{V}_{s,t} \in \mathbb{R}^{X,Y}$, where X and Y represent the vertical and horizontal dimension of each sub-aperture image. Each sub-aperture image corresponds to a fixed pair of (s, t).

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

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

with $vec(\mathbf{V}_i)$ being the vectorized representation of the $i-$th sub-aperture image, m represents the number of pixels in each view ($m = X \times Y$) and n is the number of views in the light field.

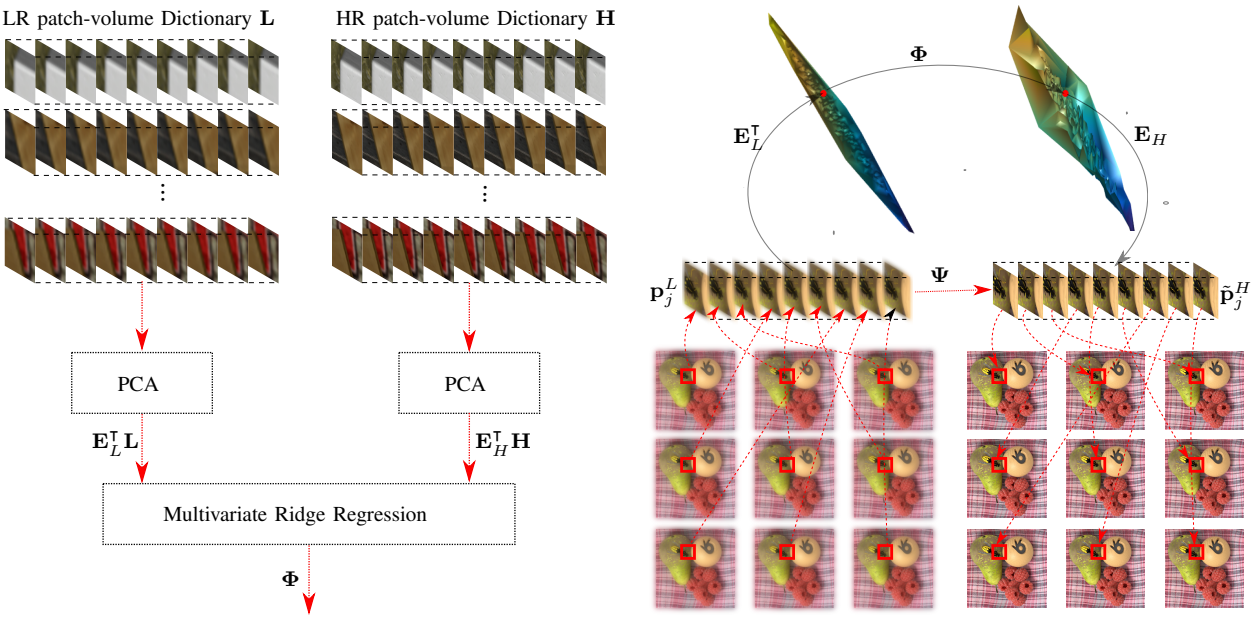
We define here a *patch-volume* $\mathbf{p}_j \in \mathbb{R}^{q,q,n}$ to be a volumetric stack of 2D patches of dimension $q \times q$ from each of the n sub-aperture images, where j stands for the patch index. Let (x_{c_j}, y_{c_j}) be the spatial coordinates of the center pixel of the $j-$th patch. The *patch-volume* $\mathbf{p}_j$ is therefore made of a stack of 2D-patches extracted from each sub-aperture image, each 2D-patch including all pixels within the range $\big([x_{c_j} - \frac{q}{2}, x_{c_j} + \frac{q}{2}], [y_{c_j} - \frac{q}{2}, y_{c_j} + \frac{q}{2}]\big)$. Without loss of generality, the *patch-volume* will be assumed to be centered (by mean removal) in the sequel.

III. Proposed Method

Let $\mathbf{V}^L$ be the low-resolution input light field matrix and $\mathbf{V}^H$ its high-resolution version which we try to estimate. Given the observed low-resolution light field $\mathbf{V}^L$, this problem can be formulated in a Banach space as

$$\mathbf{V}^L = \Theta \mathbf{V}^H + \boldsymbol{\eta} \tag{2}$$

where $\boldsymbol{\eta}$ is an additive noise matrix and Θ is a bounded operator composed of a blurring kernel $\mathbf{B}$ and a downsampling operator by a factor α applied on each sub-aperture image. There are many possible high-resolution light fields $\mathbf{V}^H$ which can produce the input low-resolution light field $\mathbf{V}^L$ via the acquisition model. Hence, solving this ill-posed inverse problem requires introducing some priors on $\mathbf{V}^H$, which can be a statistical prior such as a GMM model [8], or priors learned from training data as in [9] and in our proposed method.



(a) Training Phase (b) Testing Phase

Fig. 1. Schematic diagram of the proposed spatial Light Field Super-Resolution method where (a) shows how the linear projections are learned from a set of training *patch-volumes* and (b) how the learned projections are used to restore the Light Field. The black arrows in (b) show the actual projection of the low-quality patch volume $\mathbf{p}_j^L$ onto the low-quality sub-space using $\mathbf{E}_L^\intercal$, followed by the projection from the low- to high-quality subspaces using the projection $\mathbf{\Phi}$ and then the projection from the high-quality subspace to reconstruct the high-quality patch volume $\tilde{\mathbf{p}}_j^H$. Note that the projection $\mathbf{\Psi}$ is equal to the product of these linear projection matrices as defined in equation (8).

A. Method overview

Fig. 1 depicts the block diagram of the proposed spatial light field super-resolution algorithm, which consists of a Training Phase (Fig. 1 (a)) used to learn the linear projections and a Testing Phase (Fig. 1 (b)) where the linear projections are used to restore the light field. The Training phase depicted in Fig. 1 (a) is explained in more detail in sub-sections III-B and III-C. This sub-section will explain how the learned linear projections are used to restore a low-quality light field.

During Testing phase (see Fig. 1 (b)), each sub-aperture image of the unseen[1] low-quality light field $\mathbf{V}^L$ is first bi-cubic interpolated, so that both $\mathbf{V}^L$ and $\mathbf{V}^H$ have the same dimensions, and then divided into a set of overlapping patches of size $q \times q$ with an overlap of γ which, without loss of generality, was set to $\lfloor \frac{q}{3} \rfloor$. While the authors of the previous patch-based approach for super-resolving light fields [8] use a LMMSE approach with a GMM prior to estimate the 4D high-resolution patches, we instead propose an example-based method operating on 3D stacks of 2D-patches extracted from the different sub-aperture images. The stack of 2D-patches is called a *patch-volume* which is of dimension $z = q \times q \times n$. To form the stack of 2D-patches, in the sequel, we consider two cases: i)-stacking the 2D-patches which are at the same spatial location in all sub-aperture images, and ii)-stacking best matching 2D-patches across all sub-aperture images.

Example-based super-resolution methods for 2D images often rely on the assumption that the low-resolution (and respectively high-resolution 2D patches) lie on (or close to) a manifold. Neighbour embedding techniques, as in [18], [19], [20], were proposed for single image super-resolution. They aim at capturing the local geometry of a manifold on which the low-resolution patches are supposed to lay, by computing linear combination weights to approximate a given 2D patch from its neighbours on the manifold. The same weights are then used to reconstruct the high-resolution 2D patch. More recently, the authors in [21], [22] partition the dictionaries into multiple clusters and use a local linear mapping for each cluster to compute the weights. All these methods assume that both low- and high-resolution patches lie on two manifolds having similar geometrical structures, hence on which the patch neighborhood is preserved. In order to enforce the assumption of manifold similarity between the two distinct spaces represented by the LR and HR patches, the authors in [23] compute the weights in a subspace common to the LR and HR patches, assuming the existence of a common low-dimensional space that preserves some meaningful patch characteristics. However, the authors in [24] have shown that this assumption does not hold well because the one-to-many mapping from low- and high-resolution manifolds distort the structure of the low-resolution manifold, which negatively affects the neighbour preservation.

Here, we consider a different patch reconstruction approach, similar to the work in [15], [25], [26] which consists in learning a mapping function between the low- and high-resolution sub-spaces. Nevertheless, these existing methods are designed for single-image super-resolution and do not exploit the light field structure to restore the light field. Moreover, learning the mapping relation for *patch-volumes* is not a trivial task since each *patch-volume* has a considerably

[1]Patches from the low-quality light field are not included in the coupled dictionaries $\mathbf{L}$ and $\mathbf{H}$.

large dimension z which can be of the order of thousands of pixels. As the sub-aperture images correspond to different representations of the same scene captured from different viewpoints, these *patch-volumes* contain a lot of redundant information that can be exploited to reduce the dimension of the *patch-volumes* and therefore making the learning of the mapping between low- and high-resolution *patch-volumes* easier. Dimensionality reduction techniques can be used to project the low- and high- resolution *patch-volumes* onto their corresponding low-dimensional subspaces, and the learning of the mapping function can be performed between projected training *patch-volumes*.

In summary, in order to learn this mapping function, we first construct coupled dictionaries $\mathbf{L}$ and $\mathbf{H}$ of dimension $\mathbb{R}^{z,M}$ which respectively contain low- and high-resolution *patch-volumes* to be used for training. These dictionaries are not learned and are constructed by choosing M random *patch-volumes* from a set of light fields captured using the same camera model. The low-resolution *patch-volumes* are first projected onto the low-resolution subspace using the projection matrix $\mathbf{E}_L \in \mathbb{R}^{z,d_l}$ ($d_l \ll z$ is the dimension of the low-resolution subspace), and then projected onto the high-resolution subspace using a projection matrix $\mathbf{\Phi} \in \mathbb{R}^{d_h,d_l}$ ($d_h \ll z$ is the dimension of the high-quality subspace), and then projected back to the pixel domain using the projection matrix $\mathbf{E}_H \in \mathbb{R}^{z,d_h}$. In the next sections we detail the subspace projection, the mapping function learning steps and how this scheme can be extended to perform angular super-resolution

B. Projection on patch-volume subspaces

The vectorized input low-resolution *patch-volume* $\mathbf{p}_j^L$ and the vectorized high-resolution *patch-volume* $\mathbf{p}_j^H$ have a dimension z of the order of thousands of pixels. In order to reduce the dimension of the *patch-volumes*, PCA is adopted to compute the projection matrices for both low- and high-resolution *patch-volume* subspaces. The low-resolution *patch-volume* dictionary $\mathbf{L}$ and the high-resolution *patch-volume* dictionary $\mathbf{H}$ are used to compute respectively the low- and high-resolution *patch-volume* subspaces. The first step involves the computation of the respective covariance matrices

$$\mathbf{C}_L = \mathbf{L}\mathbf{L}^\intercal \qquad \mathbf{C}_H = \mathbf{H}\mathbf{H}^\intercal \tag{3}$$

The low-resolution *patch-volume* subspace projection matrix $\mathbf{E}_L$ and the corresponding high-resolution *patch-volume* subspace projection matrix $\mathbf{E}_H$ are respectively derived using the eigen-decomposition of the corresponding covariance matrices $\mathbf{C}_L$ and $\mathbf{C}_H$. Both low- and high-resolution *patch-volumes* can be projected on their corresponding subspaces whose dimension is significantly lower than the size z of the *patch-volume* while still keeping the most relevant information. One appealing property of these projections computed using PCA is that both projection matrices $\mathbf{E}_L$ and $\mathbf{E}_H$ are orthogonal and therefore one can project a *patch-volume* to and from the subspace without loosing any information *i.e.* $\mathbf{E}_L\mathbf{E}_L^\intercal = \mathbf{E}_L^\intercal\mathbf{E}_L = \mathbf{I}$ and $\mathbf{E}_H\mathbf{E}_H^\intercal = \mathbf{E}_H^\intercal\mathbf{E}_H = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. Nevertheless, we use PCA to reduce the dimension of the *patch-volumes* and therefore we only use the first d_L column vectors from $\mathbf{E}_L$, and the first d_H column vectors from $\mathbf{E}_H$ based on the largest eigen-values. Reducing the number of column vectors from the projection matrices has the benefit of reducing the dimension of the *patch-volume* sub-spaces at the expense of loosing some information which may not be significant. Nevertheless, reducing the dimensionality of the problem will facilitate the learning of the mapping function between the low- and high-resolution *patch-volume* subspaces.

C. Linear mapping between LR and HR subspaces

Patch-volume based super-resolution can be viewed as a regression problem *i.e.* finding a mapping function $\mathbf{\Psi} \in \mathbb{R}^{z,z}$ from the low-resolution *patch-volume* to the target high-resolution *patch-volume*. However, given that the dimension of the *patch-volume* z is generally very large, the regression learning process needs a huge amount of data ($M \gg z$) in order to learn a suitable regression model. For example, the authors in [9] employed between $1-2$ million patches to train their DCNN which takes a huge amount of time to train.

Instead, we try to learn the projection between the low- and high-resolution *patch-volume* subspaces. The major advantage of this approach is that the low- and high-resolution *patch-volume* subspaces have a dimension of d_L and d_H respectively, where both of them are significantly lower than z and therefore a smaller number of training *patch-volume* samples are needed.

We therefore use the low- and high-resolution *patch-volumes* in the coupled dictionaries, $\mathbf{L}$ and $\mathbf{H}$, which are projected respectively onto the low- and high-resolution *patch-volume* subspaces using

$$\mathbf{L}_s = \mathbf{E}_L^\intercal\mathbf{L} \qquad \mathbf{H}_s = \mathbf{E}_H^\intercal\mathbf{H} \tag{4}$$

yielding the projected coupled dictionaries $\mathbf{L}_s$ and $\mathbf{H}_s$ (see Fig. 1 (a)). The mapping function is then learned by minimizing the empirical fitting error between all the pairs of examples on the two subspaces. We consider a linear mapping function, *i.e.* we search for the projection matrix $\mathbf{\Phi}$ which minimizes the following l_2- norm regularized least squares problem

$$\mathbf{\Phi} = \arg\, min_{\mathbf{\Phi}} ||\mathbf{H}_s - \mathbf{\Phi}\mathbf{L}_s||_2^2 + \lambda||\mathbf{\Phi}||_2^2 \tag{5}$$

where λ is a regularization parameter which was set to 10^{-6} in all our experiments. The regularization term is a Tikhonov regularization added to prevent singular values and provide a more stable solution than ordinary least squares. The problem is solved using multivariate ridge regression, and the solution can be written in closed form as

$$\mathbf{\Phi} = \mathbf{H}_s\mathbf{L}_s^\intercal\left(\mathbf{L}_s\mathbf{L}_s^\intercal + \lambda\mathbf{I}\right)^{-1} \tag{6}$$

where $\mathbf{I}$ is the identity matrix. The high-resolution *patch-volume* can therefore be approximated using

$$\tilde{\mathbf{p}}_j^H = \mathbf{\Psi}\mathbf{p}_j^L \tag{7}$$

where $\mathbf{\Psi}$ is a projection matrix that can be pre-computed using

$$\mathbf{\Psi} = \mathbf{E}_H\mathbf{\Phi}\mathbf{E}_L^\intercal \tag{8}$$

The restored patch-volume $\tilde{\boldsymbol{p}}_j^H$ is de-multiplexed to restore the collocated patches within each sub-aperture image simultaneously. Overlapping regions within the *patch-volume* are averaged to minimize blocking artefacts.

The complexity to restore a *patch-volume* is of the order $O(z^2)$. In terms of execution, the proposed PCA+RR takes 158-seconds to restore the INRIA light fields while the SRCNN+LF [9] method [2] and the method of Mitra [8] are computed in 96- and 509-seconds respectively. The other methods take more than 15-minutes to restore a single INRIA light field.

D. Patch-Volume Alignment

The PCA+RR method discussed above assumes that the stack of 2D collocated patches from all sub-aperture images, which are part of the same *patch-volume*, are well aligned. However, this may not always be the case, especially when filmed using arrays of cameras capturing the scene from different viewpoints. In such systems, large disparities will exist across collocated 2D patches forming the *patch-volume*. To characterize this problem we derive a variance map $\boldsymbol{\Sigma}$ which is computed as

$$\boldsymbol{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\boldsymbol{V}_i^L - \boldsymbol{\mu}_i\right)^2 \tag{9}$$

where

$$\boldsymbol{\mu}_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{V}_i^L \tag{10}$$

The variance map $\boldsymbol{\Sigma}$ measures the variance across the n sub-aperture images. Fig. 2 shows an analysis conducted on the Buddha synthetic light field from the HCI dataset. It can be seen that the proposed PCA+RR method performs well when the pixel-variance is sufficiently low (first two rows in Fig. 2). On the other hand, the reconstructed image contains severe distortions in regions having large pixel-variance (bottom two rows in Fig. 2). Therefore, the objective here is to develop a method which reduces the variance within each *patch-volume*.

In this work, we proposed an iterative procedure where a different sub-aperture image is chosen as anchor at each iteration. The anchor sub-aperture image is used to extract overlapping 2D anchor patches. The block matching algorithm (BMA) is then used to search within a search window for the 2D patch from the other sub-aperture images which minimizes the mean square error. Therefore, the BMA method will derive those patches which are best aligned and therefore is expected to reduce the variance within the *patch-volume*.

At each iteration, all pixels within the anchor sub-aperture image are restored. However, the remaining sub-aperture images will contain pixels which have not been super-resolved (not considered when forming the *patch-volumes* by the block matching search). The sub-aperture image with the largest number of unprocessed pixels (holes) is selected as the new anchor sub-aperture image for the next iteration. This process terminates once all pixels within the light field have been processed. The BM+PCA+RR algorithm is summarized in algorithm 1 and a demo will be made available upon publication.

[2]The SRCNN+LF method was implemented using Caffe.

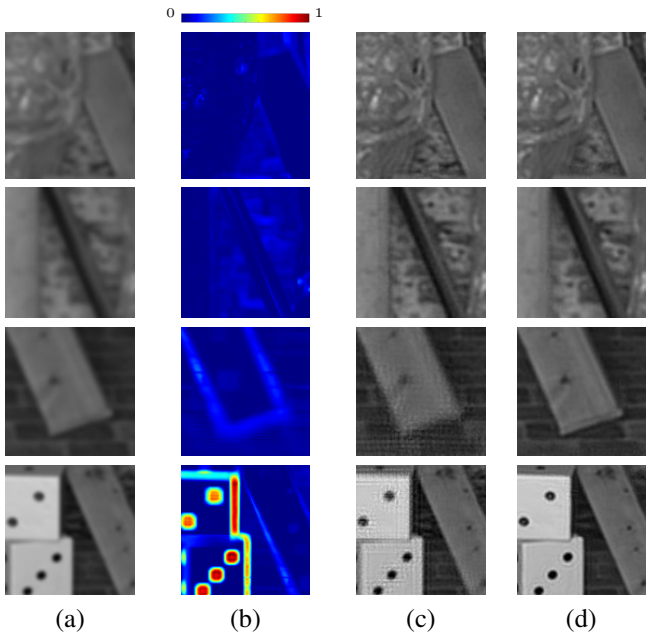


Fig. 2. Analysing different sub-aperture images from the center view of the Buddha light field (a) low-quality sub-aperture image, (b) variance map $\boldsymbol{\Sigma}$, and images super-resolved with the proposed (c) PCA+RR and (d) BM+PCA+RR methods (see Algorithm 1)

In order to characterize the variance within each *patch-volume*, we define the *patch-volume variance* $\boldsymbol{\sigma}_j$ which is computed using

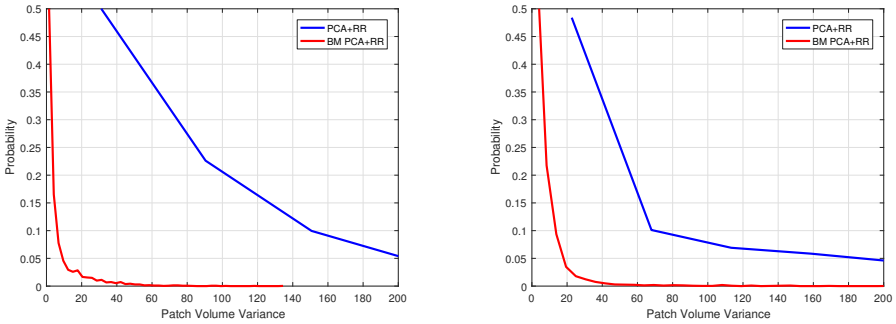
$$\boldsymbol{\sigma}_j = \frac{1}{q^2}\sum_{j}\boldsymbol{\Sigma}_j \tag{11}$$

which is the average variance within the $j-$th $q \times q$ patch from $\boldsymbol{\Sigma}$. Fig. 3 shows the distribution of $\boldsymbol{\sigma}_j$ with and without block matching for the Buddha and Still Life light fields. It can be immediately noticed that block matching manages to significantly reduce the *patch-volume variance*. It can also be seen from the last two columns of Fig. 2 that reducing the *patch-volume variance* significantly improves the quality of the super-resolved light field, especially in regions with high disparity.

E. Angular Super-Resolution

A simple way to produce a novel view between two neighbouring sub-aperture images is to apply bilinear interpolation. However, such approach loses high frequency information and produces sub-aperture images that are blurred. On the other hand, the authors in [7] assume that accurate depth maps are used as a geometric guidance to perform angular SR. However, as shown in [9], this method fails in practice because obtaining reliable depth-maps from a light field is prone to errors which significantly reduces its performance.

The DCNN based scheme in [9] exploits the local neighbourhood, but tries to approximate a global model for all



(a) (b)

Fig. 3. Probability distribution of the *patch-volume variance* $\boldsymbol{\sigma}_j$ for the (a) Buddha and (b) Stilllife light fields.

Algorithm 1: The BM+PCA+RR Light field super-resolution algorithm

$\tilde{\boldsymbol{V}}^H$ = function bm_pca_rr_LFSR $(\boldsymbol{V}^L, \boldsymbol{\Psi})$;
Initialize: Set the center view as the anchor sub-aperture image *i.e.* $i_A = \lceil \frac{n}{2} \rceil$, where n is the number of sub-aperture images
while *not all pixels of the light field are reconstructed* **do**
  for $j = 1 : P$ **do**
    Choose the $j-$th 2D patch, denoted by $\boldsymbol{p}^L_{i_A,j}$, from the anchor sub-aperture image and put it as the i_A-th patch in the *patch-volume* $\boldsymbol{p}^L_j$.
    for $i = 1 : N$ **do**
      if $i \neq i_A$ **then**
        Use the block-matching algorithm to find the patch within the $i-$th sub-aperture image which is closest to $\boldsymbol{p}^L_{i_A,j}$ in terms of mean square error and put it as the $i-$th patch in the *patch-volume* $\boldsymbol{p}^L_j$.
      end
    end
    Project the low-quality patch-volume $\boldsymbol{p}^L_j$ using (7) to synthesize the high-quality *patch-volume* $\tilde{\boldsymbol{p}}^H_j$.
    Demultiplex the reconstructed patch-volumes to reconstruct the high quality light field, where overlapping pixels are averaged.
  end
  Compute the number of holes (pixels which have not been selected by BMA) in each sub-aperture image and set the image with the largest number of holes as the new anchor sub-aperture image.
end

angular viewpoints. However, deriving the angular super-resolution of a light field using one global function generally provides poor performance. This can be explained by the fact that each sub-aperture image represents the same scene from different viewpoints, and therefore a global interpolation function suitable for all different angular displacements will provide sub-optimal performance. Instead, we employ a new method where a linear interpolation function for each missing sub-aperture image is learned as shown in Fig. 4. Similar to

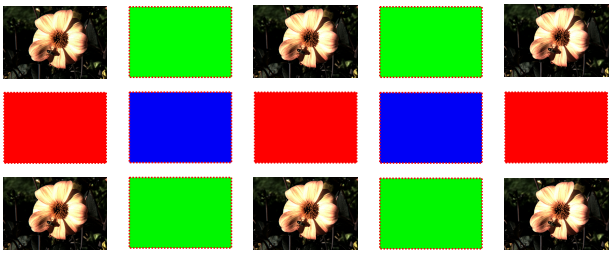


Fig. 4. Schematic diagram of the proposed Angular Light Field Super-Resolution method.The green boxes represent case 1), red boxes represent case 2) and blue boxes represent case 3). Note that the green, red and blue boxes are empty during testing phase, but these sub-aperture images are available in $\mathbf{H}$ during training

the work of Yoon *et. al.* [9], our approach is a data driven supervised learning method. However, in our case we use multivariate ridge regression to learn the interpolation function for each unknown view-point (s,t). Moreover, while one interpolation function is learned for every missing sub-aperture image, the proposed angular super-resolution algorithm is still a learning patch-based strategy. We here define three cases to learn a mapping function:

1) The case where sub-aperture images at locations $(s-1,t)$ and $(s+1,t)$ are available. In this case we extract the 2D patches at locations $(s-1,t)$, (s,t) and $(s+1,t)$ from the *patch-volumes* contained in the high-quality dictionary $\mathbf{H}$. All the 2D patches at location $(s-1,t)$ and $(s+1,t)$ are concatenated to form the dictionary of the known part $\mathbf{H}^k_i$ while the dictionary of unknown part $\mathbf{H}^u_i$ is formed using all 2D patches at location (s,t).
2) The case where sub-aperture images at locations $(s,t-1)$ and $(s,t+1)$ are available is very similar to the above case. This time the 2D patches from $\mathbf{H}$ at location $(s,t-1)$ and $(s,t+1)$ are concatenated to form the dictionary of known part $\mathbf{H}^k_i$ while the 2D patches at location (s,t) are used to construct the dictionary of unknown part $\mathbf{H}^u_i$.
3) The case where the neighbouring horizontal and vertical sub-aperture images are missing. In this case, the 2D patches from $\mathbf{H}$ at locations $(s-1,t-1)$, $(s-1,t+1)$,$(s+1,t-1)$ and $(s+1,t+1)$ are concatenated to construct the dictionary of the known part $\mathbf{H}^k_i$ while the

2D patches at location (s,t) are used to construct the dictionary of unknown part $\mathbf{H}_i^u$.

The projection matrix $\boldsymbol{\beta}_i$ is computed for every unknown sub-aperture image i, and is used to estimate those sub-aperture images which need to be interpolated. Therefore, for a 5×5 matrix of sub-aperture images, we need 56 $\boldsymbol{\beta}_i$ projection matrices (one for every missing sub-aperture image) to approximate the missing sub-aperture images to increase the angular resolution to 9×9. Therefore, unlike the DCNN scheme, our method models the angular super-resolution problem using multi-linear models, one for every missing view. The projection matrix is approximated using multivariate ridge regression using its closed form solution

$$\boldsymbol{\beta}_i = \mathbf{H}_i^u \mathbf{H}_i^{k\intercal} \left(\mathbf{H}_i^u \mathbf{H}_i^{u\intercal} + \lambda \mathbf{I}\right)^{-1} \tag{12}$$

where λ is a regularization parameter set to 10^{-6}, and $\boldsymbol{\beta}_i$ is the interpolation matrix for the $i-$th unknown sub-aperture image. The projection matrix for each unknown sub-aperture image $\boldsymbol{\beta}_i$ can be pre-computed.

During the testing phase, the input light field is first up-scaled in the angular dimension by inserting empty sub-aperture images that need to be restored using the proposed method. Therefore, each *patch-volume* $\mathbf{p}_j$ contains a number of empty 2D patches which need to be approximated. The neighbourhood of the empty 2D patches $\mathbf{p}_j(s,t)$ within the *patch-volume* are used to determine which of the above mentioned cases will be used to construct the vector $\boldsymbol{\Gamma}_j$. In case the horizontal neighbours are available, $\boldsymbol{\Gamma}_j$ is constructed by concatenating the vectorized representations of the 2D patches $\mathbf{p}_j(s-1,t)$ and $\mathbf{p}_j(s+1,t)$. Otherwise, if the vertical neighbours are available, $\boldsymbol{\Gamma}_j$ is constructed by concatenating the vectorized representations of the 2D patches $\mathbf{p}_j(s,t-1)$ and $\mathbf{p}_j(s,t+1)$. The last case is when neither of the vertical and horizontal neighbours are available, in which case the vectorized representation of $\mathbf{p}_j(s-1,t-1)$, $\mathbf{p}_j(s-1,t+1)$, $\mathbf{p}_j(s+1,t-1)$ and $\mathbf{p}_j(s+1,t+1)$ are concatenated to construct $\boldsymbol{\Gamma}_j$. The unknown patch is then approximated using

$$\tilde{\mathbf{p}}_j(s,t) = \boldsymbol{\beta}_i \boldsymbol{\Gamma}_j \tag{13}$$

In this work we only consider collocated patches since only local neighbouring 2D patches are used, where the disparities are expected to be relatively low.

IV. Experimental Results

The experiments conducted in this paper use both synthetic and real-world light fields from publicly available datasets. Here we use the synthetic light fields of HCI[3] for parameter selection and also to get a better understanding of the performance of the proposed method. It was also tested on two real-world light fields captured using different technologies: i) the Stanford dataset[4] captured using a moving camera mounted on a gantry, where the captured light fields generally have larger disparities across sub-aperture images and ii) the INRIA dataset[5] which was captured using micro-lens technology (a Lytro Illum camera) with lower disparities across sup-aperture images. While the sub-aperture images of the Stanford data set are available, the light fields from the INRIA database were decoded using the method in [27] as mentioned in the website.

In all the experiments we considered a 9×9 array of sub-aperture images from each dataset. For computational purposes, the high resolution images of the synthetic HCI light fields were downscaled to 384×384 per sub-aperture image. The Stanford light fields have different resolutions, but the high resolution sub-aperture images were scaled such that the lowest dimension was set to 400 pixels while maintaining the original aspect ratio. On the other hand, the original light field resolutions were used for the INRIA dataset *i.e.* 625×434. In all the experiments involving spatial super-resolution, each sub-aperture image was blurred using a Gaussian filter of size 7×7 with standard deviation of 1.6 followed by downscaling by a factor of α, where unless otherwise specified is set to $\alpha = 2$.

In every experiment we use a leave-one out methodology to train the projection matrix $\boldsymbol{\Psi}$ for every dataset considered *i.e.* if we take the Buddha light field from the HCI dataset as an example, we extract 2000 *patch-volumes* from all the other light fields of the same dataset except for the Buddha. This was done to learn the mapping function $\boldsymbol{\Psi}$ for the specific camera model, without of course biasing the results by including *patch-volumes* from the light field under test. The total number of patch-volumes varied depending on the number of light fields considered for each dataset: i) 16K for the HCI dataset ii) 22K for the Stanford dataset iii) 14K for the INRIA dataset. While we observed that better performance can be achieved using a larger set of training *patch-volumes*, it did not significantly improve when using dictionaries larger than the above mentioned sizes.

The performance of the proposed method was compared to the single-image super-resolution schemes [13], [14], [15], where these methods are applied on every sub-aperture image. It was also compared against two light-field super-resolution methods including the work of Mitra *et. al.* [8] and the DCNN based scheme [9]. The angular super-resolution method was also compared against the recent view synthesis method published in [11]. In all the experiments we used the code and pre-trained models provided by the authors or are publicly available on-line.

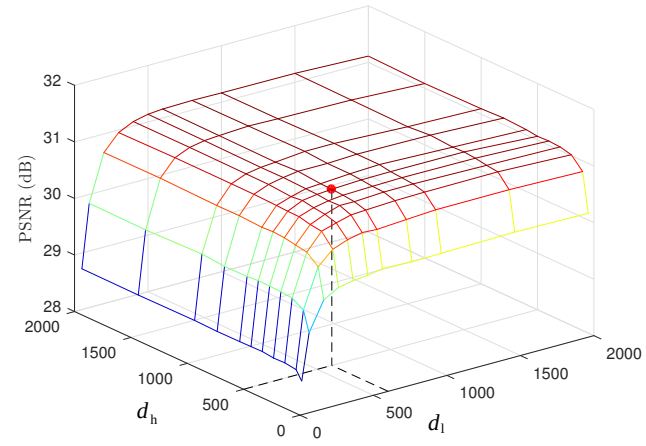
A. Parameter Selection

The method presented in this paper has four parameters that have to be tuned: i) the dimension of the low-resolution *patch-volume* subspace d_l, ii) the dimension of the high-resolution *patch-volume* subspace d_h, and iii) the patch size q and iv) the number of *patch-volumes*. Fig. 5 shows the performance in terms of PSNR using the PCA+RR method obtained by fixing the patch size q and varying the dimension of the subspaces. Similar plots were obtained using different light fields and different patch sizes but due to space constraints could not be

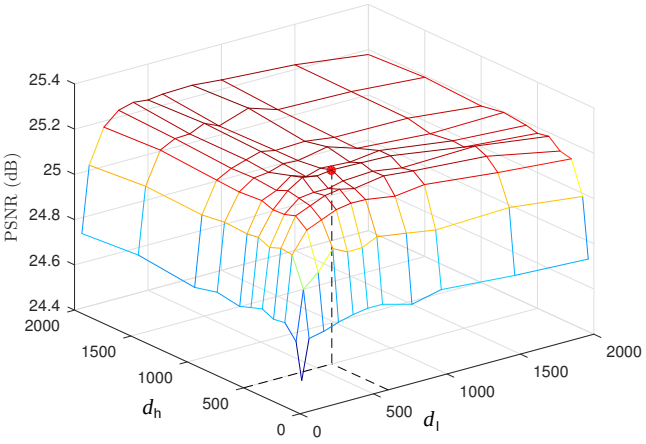
[3]HCI dataset: https://hci.iwr.uni-heidelberg.de/hci/softwares/light_field_analysis

[4]Stanford dataset: http://lightfield.stanford.edu/

[5]INRIA dataset: http://www.irisa.fr/temics/demos/lightField/CLIM/DataSoftware.html

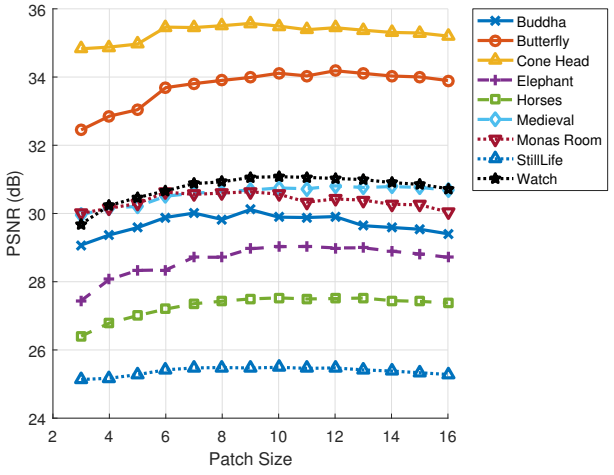


Watch Light Field with $q = 9$

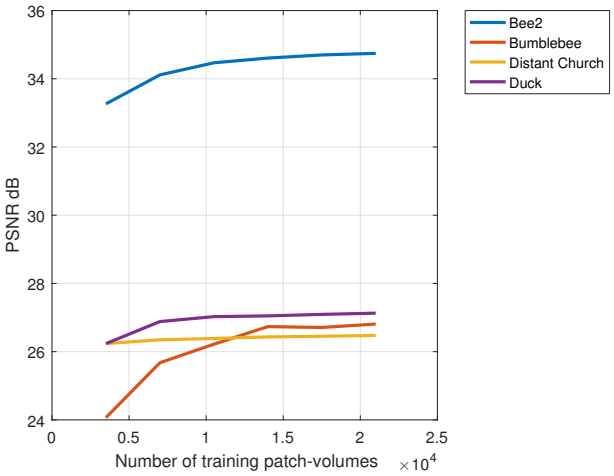


Still Life Field with $q = 5$

Fig. 5. PSNR analysis computed by setting using different light fields from the HCI dataset and different patch sizes.



(a)



(b)

Fig. 6. PSNR analysis as a function of (a) patch size and (b) number of training *patch-volumes* for different light fields

included in this paper. It can be seen that for different patch sizes and light fields the graphs seem to saturate and that good performance is achieved when setting $d_l = d_h = 500$.

The optimal patch size q is derived using the PSNR analysis illustrated in Fig. 6 (a) where we set $d_l = d_h = 500$ and use nine light fields from the HCI dataset. It can be seen that the performance starts increasing with increasing patch sizes and it saturates (or starts degrading) at $q > 10$. Therefore setting $q = 10$ seems to be an optimal parameter for the synthetic light fields. Unless otherwise specified, we will use these parameters for the following simulations. In order to assess the effect of the number of training *patch-volumes* we super-resolve four INRIA light fields with a magnification factor of 2. The results in Fig. 6 (b) show that higher PSNR is achieved when using more training *patch-volumes*. Nevertheless, the performance does not significantly improve when using more than 14K *patch-volumes* for training.

B. Spatial Super-Resolution of Synthetic Light Fields

In this section we analyse the performance of the proposed method using the synthetic light fields of the HCI dataset. In these experiments we set the magnification factor $\alpha = 2$ and compare our approach to state-of-the-art methods in the area of single-image super-resolution [13], [14], [15] and light field super-resolution [8], [9]. The results in Table I and Fig. 7 clearly show the superiority of the proposed method where it achieves PSNR gains of up to 4dB over the second best algorithm. It can be seen, especially from the images in Fig. 7, that single-image super-resolution schemes and bi-cubic interpolation generally provide blurred sub-aperture images. On the other hand, both the recent deep learning based scheme and our proposed method generate images which are sharper, with our proposed BM+PCA+RR providing the sharpest and most accurate results. Unlike existing schemes, our proposed method exploits the redundant information present within the light-field structure through the use of *patch-volumes* and low-dimensional sub-space projections. Moreover, this gain

TABLE I
PSNR ANALYSIS OF SPATIAL SUPER-RESOLUTION ON THE HCI DATASET

Light Field Name	Bicubic	Dong [14]	Peleg [13]	Zhang [15]	Mitra [8]	SRCNN+LF [9]	PCA+RR	BM PCA+RR
Buddha	27.2387	27.1295	27.5482	27.6509	22.2204	27.3565	29.9276	**31.1277**
Butterfly	29.7868	29.5468	30.0613	30.1555	28.8319	29.2139	**33.7600**	33.5527
Cone Head	34.2545	34.1121	34.3921	34.5577	29.5444	31.3876	35.3642	**36.6109**
Horses	24.5134	24.6663	24.8337	24.8801	23.9871	24.6014	**27.9206**	27.5649
Medieval	29.1129	29.0211	29.2461	29.3695	28.3404	28.6737	**31.1431**	31.0744
Monas Room	27.3276	27.1595	27.6295	27.7423	23.5765	27.1786	30.2960	**30.7295**
Elephant	24.4555	24.5238	24.7661	24.7881	22.8509	24.1461	**29.5462**	28.3563
Watch	25.9329	25.6987	26.2238	26.2930	23.9491	25.5008	**30.9749**	29.9094
StillLife	24.2778	24.0777	24.4509	24.5284	21.4812	23.8061	26.0804	**26.1894**

TABLE II
PSNR ANALYSIS OF SPATIAL SUPER-RESOLUTION (MAGNIFICATION FACTOR $\times 2$) USING THE STANFORD LIGHT FIELD IMAGES

Light Field Name	Bicubic	Dong [14]	Peleg [13]	Zhang [15]	Mitra [8]	SRCNN+LF [9]	PCA+RR	BM PCA+RR
Amethyst	30.0753	29.8729	30.3426	30.4042	27.5084	30.1516	33.2770	**33.3482**
Bracelet	26.2049	26.1827	26.5485	26.5931	19.0219	25.4509	25.1423	**28.4921**
Bunny	31.2884	30.8454	31.5553	31.6312	27.4327	29.4336	34.8505	**36.8623**
Chess	29.8041	29.4868	30.1000	30.1758	24.8064	29.0012	32.9157	**33.4924**
Eucalyptus Flowers	30.4913	30.3423	30.7207	30.7922	29.1275	30.5894	33.0552	**33.1837**
Jelly Beens	40.2994	39.2555	40.4947	**40.6129**	26.9460	29.7700	40.5246	38.8378
Lego Bulldozer	26.0729	25.8941	26.4372	26.4999	20.2748	26.4719	26.8637	**29.0056**
Lego Gantry Self Portrait	27.1717	27.0497	27.5291	27.6457	20.3284	27.4146	28.5695	**28.8770**
Lego Knights	26.8735	26.7013	27.2120	27.2896	19.7672	26.0193	28.0270	**29.6041**
Lego Truck	30.4957	30.3576	30.7638	30.8524	28.6439	30.8273	**33.6306**	33.5345
Treasure	25.8466	25.7149	26.0605	26.1300	21.8995	26.4346	27.6640	**28.2319**

over DCNN is achieves using a significantly smaller amount of training data which can be efficiently trained and pre-computed using a normal computer without the need of GPUs to speed up the learning process. These results also show that PCA+RR performs best with light fields like Watch and Butterfly which have low disparity while BM+PCA+RR performs better with light fields having larger disparities such as Buddha and Stilllife. Moreover, unlike existing schemes, our method ensures that the reconstructed sub-aperture images are coherent across all sub-aperture images by using of *patch-volumes* and therefore exploiting the light field structure.

C. Spatial Super-Resolution of Real World Light Fields

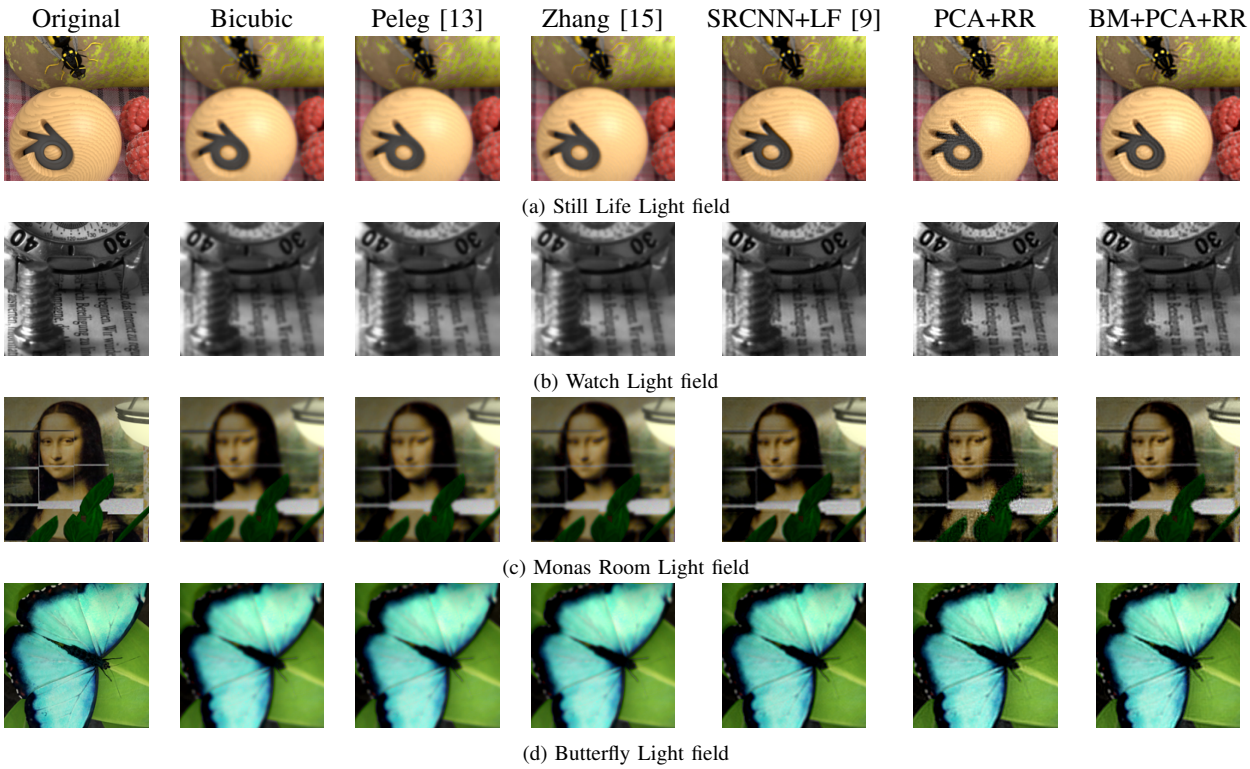
In this experiment we use the light field images provided by Stanford which are captured using a camera mounted on a gantry. The results in Table II clearly show that most of the time it is the BM+PCA+RR method which outperforms all other schemes including PCA+RR. This is mainly attributed to the fact that some of the Stanford light fields have large disparities, and the alignment using block-matching improves the performance of the reconstructed light fields by reducing the inter-patch variance.

Table III shows the average PSNR of the INRIA light fields using different super-resolution techniques. It can be seen that the proposed methods again outperform all the state-of-the-art methods considered in this work with PCA+RR having the best overall performance. This might be attributed to the fact that the light fields captured with the micro-lens technology have smaller disparities and therefore alignment of the *patch-volumes* does not improve performance. It can be mentioned at this stage that the performance of Mitra *et. al* [8], is significantly lower than the bi-cubic interpolated light fields. This confirms the results presented in [9] where similar poor performance were shown for both methods in [7], [8] which exploit depth information to perform light field super-resolution. The results in Table IV further demonstrate the performance of both proposed schemes at larger magnification factors ($\times 3$,$\times 4$).

The subjective evaluation in Fig. 8 shows the results on real-world light fields. In this experiment we converted each image into the YC_bC_r colorspace, where the luminance component is restored using the following algorithms while the chrominance channels were up-scaled using bicubic interpolation. Similar to the above results, the method of Peleg *et. al*. [13] provides blurred sub-aperture images. The DCNN based method in [9] gives improved performance over single image super-resolution techniques. Nevertheless, our proposed BM+PCA+RR method yields sub-aperture images with better texture detail.

D. Angular Super-Resolution

In these experiments we reduce the angular resolution by removing even column and row sub-aperture images, resulting in a matrix of 5×5 high-resolution and undistorted sub-aperture images. Angular super-resolution methods try to approximate novel sub-aperture images to increase the angular resolution of the light field. We compare our proposed angular super-resolution method with bilinear interpolation between neighbouring sub-aperture images, the SRCNN+LF method in [9] and the view synthesis method presented in [11].



(a) Still Life Light field

(b) Watch Light field

(c) Monas Room Light field

(d) Butterfly Light field

Fig. 7. Cropped regions of the central view obtained using different spatial super-resolution techniques obtained using the synthetic HCI dataset

TABLE III
PSNR ANALYSIS OF SPATIAL SUPER-RESOLUTION (MAGNIFICATION FACTOR $\times 2$) USING THE INRIA LIGHT FIELD IMAGES

Light Field Name	Bicubic	Dong [14]	Peleg [13]	Zhang [15]	Mitra [8]	SRCNN+LF [9]	PCA+RR	BM+PCA+RR
Bee2	30.5666	30.3450	30.9149	30.3152	25.1065	29.8795	**34.6064**	33.5067
Bumblebee	26.6509	26.4444	27.0462	26.4769	17.5138	27.6928	26.7346	**28.7119**
DistantChurch	24.3331	24.3298	24.5979	24.4754	24.3143	25.1071	26.4316	**26.4920**
Duck	23.5316	23.5510	23.8874	23.7356	21.4510	24.8502	**27.0484**	26.3975
Framed	27.5230	27.3720	27.9161	27.7436	25.4695	27.9237	**30.8640**	30.4825
Fruits	28.4813	28.4310	28.8230	28.6050	22.7371	29.5817	31.6108	**32.0372**
Mini	27.6342	27.6546	27.9222	27.7356	25.8211	28.2645	**30.8350**	30.2396
Rose	33.6350	33.6134	33.8568	33.8966	28.8269	34.4949	**37.4667**	36.9080

The angularly super-resolved views $(6, 6)$ are shown in Fig. 9, which corresponds to case 3 described in Section III-E. It can be noticed that the images produced using simple bilinear interpolation are generally blurred, especially when considering the synthetic light fields and the light fields provided by Stanford which have larger disparities and thus harder to approximate using this simple method. The view synthesis method presented in [11] manages to restore the light fields captured by the Illum camera relatively well, however it fails to restore the other light fields (HCI and Stanford), mainly because the model has been learned on Illum light fields and should be retrained to test the method on other types of light fields. The other deep learning method of Yoon *et. al.* [9] manages to recover some texture detail. However, the reconstructed interpolated sub-aperture images are generally noisy. Our proposed angular super-resolution approach provides sharper images closer to the ground truth and these improvements are more visible in synthetic and on images obtained from the Stanford dataset.

V. SPATIAL-ANGULAR SUPER-RESOLUTION

In the final experiment we analyse the combined spatial and angular super-resolution and compare it with the deep learning method presented in [9]. The 9×9 matrix of sub-aperture images is first angularly down-sampled by removing even rows and columns and then down-scaled by a factor of 2. Fig. 10 compares subjectively the results obtained using the state-of-the-art deep learning based method proposed in [9] with our combined scheme (spatial super-resolution using BM+PCA+RR and using the proposed angular super-resolution (Section III-E) method to restore the original angular resolution). One can notice that our proposed method manages to restore higher quality sub-aperture images both spatially (at angular resolution (5,5) which is available in the 5×5 matrix) and angularly (6,6) (which is angularly super-resolved). These results, and all results presented in the previous sections demonstrate that by properly exploiting the light field structure, linear models can effectively outperform more complicated deep learning based approaches for both angular

TABLE IV
PSNR ANALYSIS OF SPATIAL SUPER-RESOLUTION USING THE INRIA LIGHT FIELD IMAGES AT DIFFERENT MAGNIFICATION FACTORS

Light Field Name	×3					×4				
	Bicubic	ANR [28]	NCSR [29]	PCA+RR	BM+PCA+RR	Bicubic	ANR [28]	NCSR [29]	PCA+RR	BM+PCA+RR
Bee2	29.6867	30.6162	30.5056	**31.7366**	31.2349	28.4433	**29.2275**	28.3437	29.1141	28.9283
Bumblebee	25.5401	26.4189	24.7068	22.8932	**26.4321**	24.1508	**24.7748**	21.7147	20.4922	24.0889
DistantChurch	23.7253	24.1354	24.2603	24.5221	**24.5621**	22.9630	23.1522	22.5218	**23.1835**	23.1028
Duck	23.0798	23.7197	23.6979	**24.1611**	23.9780	22.1628	**22.5774**	20.4905	22.0103	21.9914
Framed	26.7072	27.4263	27.3974	28.1060	**28.1600**	25.5819	26.0354	25.8239	26.1740	**26.1741**
Fruits	27.7360	28.4712	27.9654	28.6624	**29.3045**	26.6765	**27.1361**	25.5000	26.3563	26.9609
Mini	26.9273	27.4880	27.0742	**28.8523**	28.4511	26.0561	26.3788	26.4890	**27.1800**	26.8443
Rose	32.6970	33.3592	33.0670	**34.8321**	34.5860	31.538	31.8898	31.0761	**32.6694**	32.5134

and spatial super-resolution. Supplementary multimedia files uploaded on ScholarOne show that the super-resolved light fields (spatial and angular SR) using our proposed method are more coherent across sub-aperture images for real-world light fields (less flickering across sub-aperture images) and provides superior re-focused images when compared to the DCNN scheme presented in [9].

VI. CONCLUSION

In this paper, we have proposed a novel super-resolution algorithm for light fields. The method is shown to outperform recent techniques based on deep convolutional neural networks for both spatial and angular super-resolution. In addition, the approach is simple, and does not require millions of training samples to learn the projections between high- and low-resolution subspaces and to increase the angular resolution in contrast to prior solutions based on deep convolutional networks. Experimental results show significant improvement over the state-of-the-art spatial super-resolution schemes, achieving average PSNR gains of 3.7dB, 4.2dB and 2.1dB on the HCI, Stanford and INRIA datasets respectively. Moreover, subjective results clearly demonstrate that the proposed spatial and angular super-resolution method outperforms more complex schemes based on deep learning. Future work will involve to learn dictionaries of *patch-volumes* that are more representative and to learn multiple linear mappings.

VII. ACKNOWLEDGEMENTS

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

REFERENCES

[1] R. Ng, "Digital light field photography," Ph.D. dissertation, Stanford University, Stanford, CA, USA, 2006.

[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.

[3] F. Nava and J. Luke, "Simultaneous estimation of super-resolved depth and all-in-focus images from a plenoptic camera," in *3DTVC Conference*, 2009.

[4] D. Cho, M. Lee, S. Kim, and Y.-W. Tai, "Modeling the calibration pipeline of the lytro camera for high quality light-field image reconstruction," in *ICCV*. IEEE, 2013, pp. 3280–3287.

[5] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *CVPR*. IEEE, 2008, pp. 3280–3287.

[6] T. E. Bishop, S. Zanetti, and P. Favaro, "Light field superresolution," in *ICCP*. IEEE, 2009, pp. 1–9.

[7] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *PAMI*, vol. 36, no. 3, pp. 606–619, 2014.

[8] K. Mitra and A. Veeraraghavan, "Light field denoising, light field superresolution and stereo camera based refocusing using a gmm light field patch prior," in *CVPR Workshop on Computational Cameras and Displays*, 2012.

[9] 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 International Conference on Computer Vision Workshop*, Dec 2015, pp. 57–65.

[10] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014, pp. 184–199. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-10593-2_13

[11] N. K. Kalantari, T.-C. Wang, and R. Ramamoorthi, "Learning-based view synthesis for light field cameras," *ACM Trans. Graph.*, vol. 35, no. 6, pp. 193:1–193:10, Nov. 2016. [Online]. Available: http://doi.acm.org/10.1145/2980179.2980251

[12] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive light field photography using overcomplete dictionaries and optimized projections," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 46:1–46:12, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2461912.2461914

[13] 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.

[14] 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.

[15] 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.

[16] M. Levoy and P. Hanrahan, "Light field rendering," in *Annual Conference on Computer Graphics and Interactive Techniques*, 1996, pp. 31–42. [Online]. Available: http://doi.acm.org/10.1145/237170.237199

[17] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," in *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

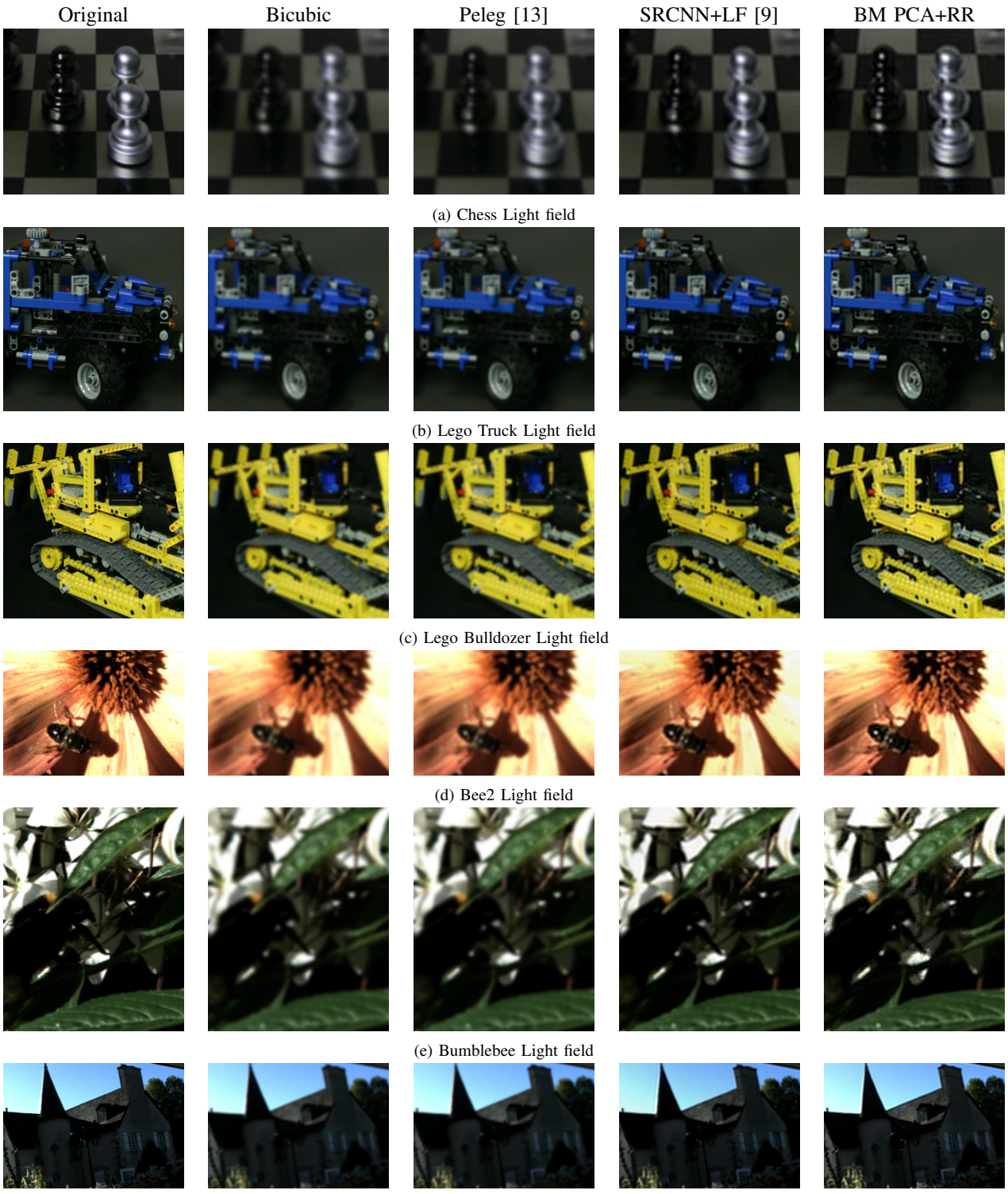
[18] H. Chang, D. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *CVPR*. IEEE, Jun. 2004, pp. 275–282.

[19] W. Fan and D. Yeung, "Image hallucination using neighbor embedding over visual primitive manifolds," in *CVPR*. IEEE, Jun. 2007, pp. 1–7.

[20] T. Chan, J. Zhang, J. Pu, and H. Huang, "Neighbor embedding based super-resolution algorithm through edge detection and feature selection," *Pattern Recognit. Lett.*, vol. 30, no. 5, pp. 494–502, Apr. 2009.

[21] K. Zhang, D. Tao, X. Gao, X. Li, and Z. Xiong, "Learning multiple linear mappings for efficient single image super-resolution," *IEEE Trans. on Image Processing*, vol. 24, no. 3, pp. 846–861, March 2015.

[22] J. C. Ferreira, E. Vural, and C. Guillemot, "Geometry-aware neighborhood search for learning local models for image superresolution," *IEEE*



(a) Chess Light field

(b) Lego Truck Light field

(c) Lego Bulldozer Light field

(d) Bee2 Light field

(e) Bumblebee Light field

(f) Framed Light field

Fig. 8. Central views spatially super-resolved using different methods on the real light fields obtained from the Stanford (a-c) and INRIA (d-f) datasets

Transactions on Image Processing, vol. 25, no. 3, pp. 1354–1367, March 2016.

[23] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. on Image Process.*, vol. 21, no. 2, pp. 469–480, Feb. 2012.

[24] K. Su, Q. Tian, Q. Xue, N. Sebe, and J. Ma, "Neighborhood issue in single-frame image super-resolution," in *2005 IEEE International Conference on Multimedia and Expo*, July 2005, pp. 4 pp.–.

[25] J. S. Choi and M. Kim, "Single image super-resolution using global regression based on multiple local linear mappings," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1300–1314, March 2017.

[26] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. A. Morel, "Single-image super-resolution via linear mapping of interpolated self-examples," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5334–5347, Dec 2014.

[27] D. G. Dansereau, O. Pizarro, and S. B. Williams, "Decoding, calibration

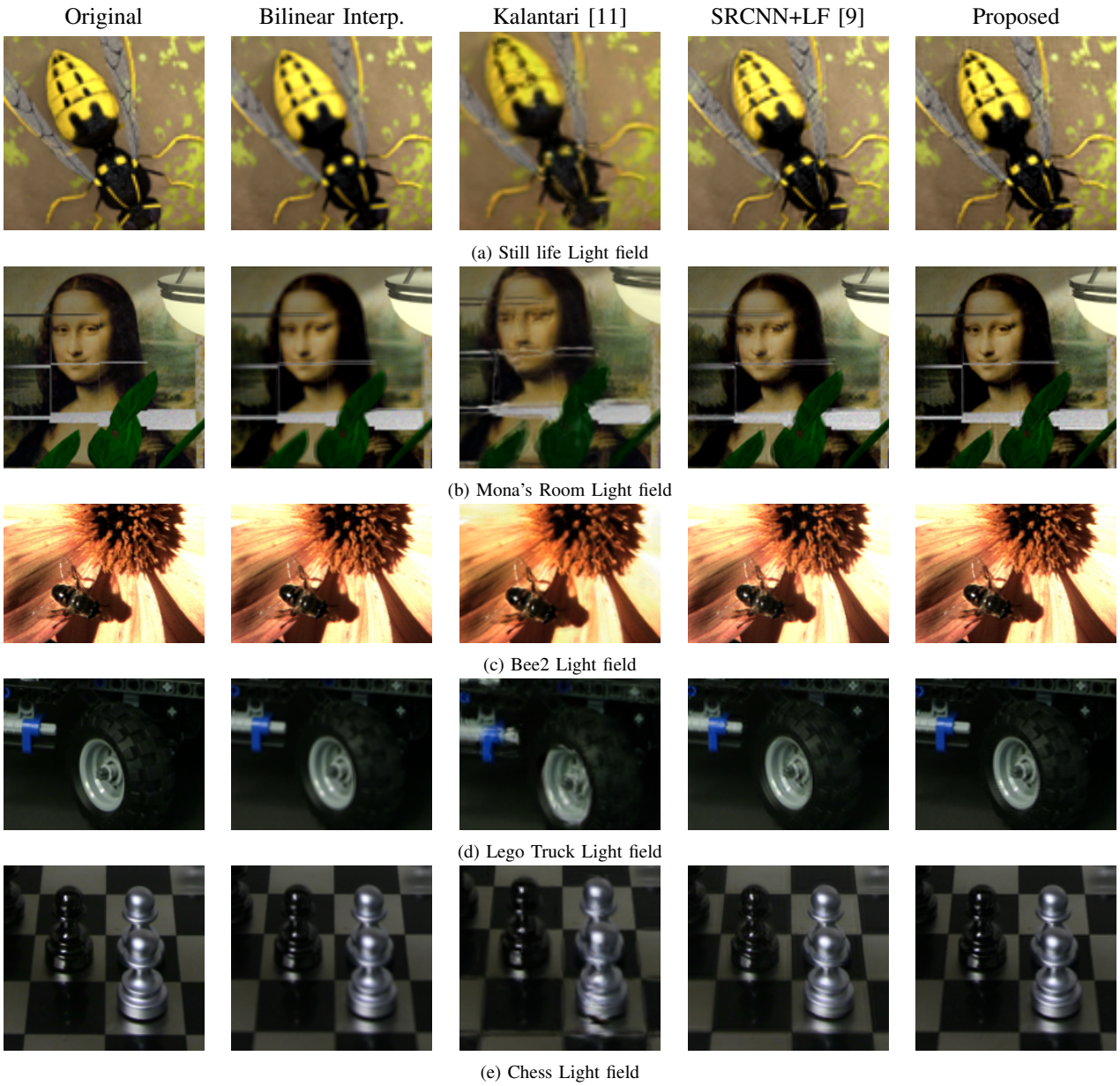


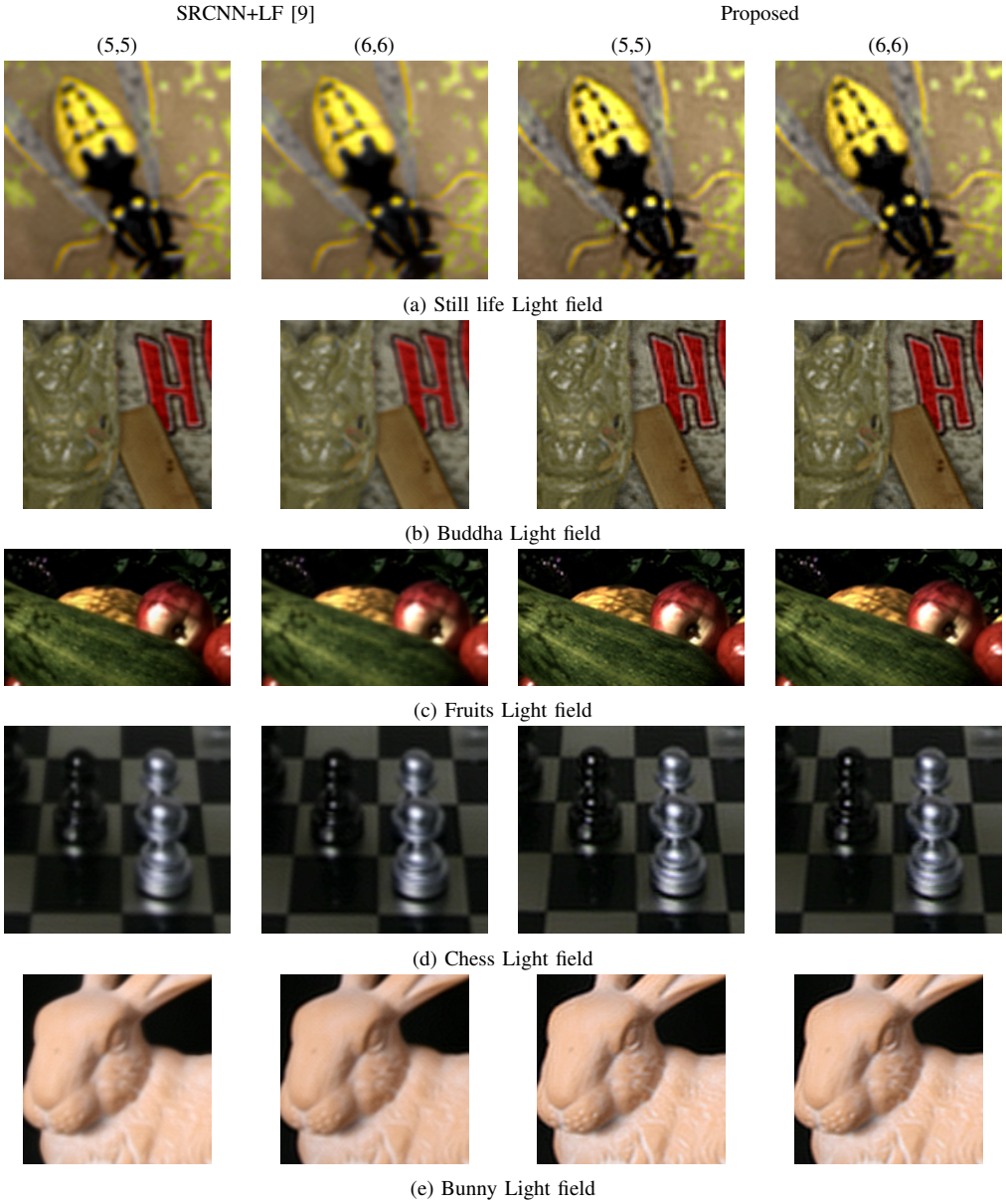
Fig. 9. Angular super-resolution analysis at interpolated sub-aperture image at (6,6) which has neither horizontal nor vertical sub-aperture neighbours. These images are obtained using the HCI (a,b), INRIA (c) and Stanford (d,e) datasets

and rectification for lenselet-based plenoptic cameras," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, June 2013, pp. 1027–1034.

[28] 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.

[29] W. Dong, L. Zhang, G. Shi, and X. Li, "Nonlocally centralized sparse representation for image restoration," *IEEE Transactions on Image Processing*, vol. 22, no. 4, pp. 1620–1630, April 2013.

Reuben A. Farrugia (S04, M09) 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).



(a) Still life Light field

(b) Buddha Light field

(c) Fruits Light field

(d) Chess Light field

(e) Bunny Light field

Fig. 10. Spatial and angular super-resolution analysis at spatially super-resolved center view (5,5) and interpolated sub-aperture image at (6,6) which has neither horizontal nor vertical sub-aperture neighbours. These images are obtained using the HCI (a,b), INRIA (c) and Stanford (d,e) datasets

Christian Gales (S'14) received the first degree in Communications and Computer Engineering in 2012 and a Masters degree in Telecommunications in 2014 from the University of Malta. He is currently pursuing a Ph.D. degree with the Faculty of Information and Communication Technology at the University of Malta. His research interests are in computer vision and image and video processing, especially applications related to objective image/video quality assessment, biometrics and automobiles.

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.