Light Field Denoising, Light Field Superresolution and Stereo Camera Based Refocussing using a GMM Light Field Patch Prior

Kaushik Mitra and Ashok Veeraraghavan
ECE, Rice University
Houston, Tx 77005
Kaushik.Mitra@rice.edu, vashok@rice.edu

Abstract

With the recent availability of commercial light field cameras, we can foresee a future in which light field signals will be as common place as images. Hence, there is an imminent need to address the problem of light field processing. We provide a common framework for addressing many of the light field processing tasks, such as denoising, angular and spatial super-resolution, etc. (in essence, all processing tasks whose observation models are linear). We propose a patch based approach, where we model the light field patches using a Gaussian mixture model (GMM). We use the "disparity pattern" of the light field data to design the patch prior. We show that the light field patches with the same disparity value (i.e., at the same depth from the focal plane) lie on a low-dimensional subspace and that the dimensionality of such subspaces varies quadratically with the disparity value. We then model the patches as Gaussian random variables conditioned on its disparity value, thus, effectively leading to a GMM model. During inference, we first find the disparity value of a patch by a fast subspace projection technique and then reconstruct it using the LMMSE algorithm. With this prior and inference algorithm, we show that we can perform many different processing tasks under a common framework.

1. Introduction

Light field is a function that describes the amount of light passing through every direction and all the points in the 3-D space. It is thus a five-dimensional function $f(x, y, z, \theta, \phi)$ with three spatial dimensions (x, y, z) and two angular dimensions (θ, ϕ). However, in free space the radiance is preserved along the direction of light travel and hence the light field can be represented by a four-dimensional function $f(x, y, \theta, \phi)$. Another popular way of representing the light field signal is the two plane parameterization $f(x, y, u, v)$ [12], which is the representation we use in this paper. The advantage of capturing the light field signal as opposed to 2-D images is that we can perform efficient novel view synthesis and refocus images after capture. To capture the light field signal many camera architectures have been proposed such as: 1) microlens array based design [16], 2) mask based design [17], 3) coded aperture based design [13] and 4) camera array based design [18]. Among these the microlens array light field cameras are comercially available as Lytro and Raytrix. Given the additional functionalities of the light field camera over the traditional cameras, we can expect that in future light field data will be as common as images are today. Hence, there is an urgent need to design efficient and robust light field processing algorithms, which can perform operations such as denoising, spatial and angular super-resolution, and other such processing tasks. In this paper, we provide a common framework for all processing tasks whose observation model is a linear function of the light field data.

Since, for many of the processing tasks, such as spatial and angular super-resolution, the number of observations is less than the number of unknown parameters (underdetermined systems), we need good prior models for the light field signal. We propose a Gaussian mixture model (GMM) prior for light field patches. Light field for a diffuse scene has a certain "disparity pattern" as shown in the X-U and Y-V slices of a light field data, see figure 1. This disparity pattern depends on the depth of the scene from the plane of focus. Consider a light field patch of dimension $n_1 \times n_2 \times s_1 \times s_2$ that correspond to a depth equal to the focal plane. All the $X-Y$ slices (there are $s_1 s_2$ such slices) of this patch are the same and hence the intrinsic or true dimension of this patch is $n_1 n_2$ (though the ambient dimension is $n_1 n_2 s_1 s_2$). From the above argument it is clear that the intrinsic dimension of light field patches depend on their corresponding depths or disparity values and that patches with the same disparity value lie on a low-dimensional subspace. We further show that the dimensionality of these subspaces are quadratically proportional to the disparity values. We then learn different Gaussian patch priors for a discrete set

978-1-4673-1612-5/12/$31.00 ©2012 IEEE

of disparity values. Assuming a uniform prior on the disparity values, our overall patch prior is thus a GMM. Given a particular processing task, which is characterized by it's observation model, we propose a fast inference algorithm. First we extract patches from the observed data,then we estimate their disparity values using a fast subspace projection algorithm and finally we estimate the corresponding light field patches using 'linear minimum mean square estimator' (LMMSE) algorithm [8]. We demonstrate the effectiveness of the proposed approach for solving many light field processing tasks such as denoising, angular and spatial light field super-resolution.

To summarize, the technical contributions of this paper are:

- We propose a common framework for solving many different light field processing tasks using a GMM prior for light field patches.
- We show that patches with a common disparity value lie on a low-dimensional subspace. We compute the dimensionality of these subspaces for different disparity values, which provide us an estimate of the minimum number of observations needed to reliably reconstruct light field patches of different disparity values.
- We provide an efficient algorithm for solving the inference problem. We use a fast subspace projection algorithm to estimate the disparity value of an observed patch and then use the LMMSE algorithm to perform the final reconstruction.

2. Prior Work

Light field prior has been used implicitly in many of early light field works such as [12, 6, 18, 16, 1]. The priors that have been used most frequently are based on the assumption that light field data are low-frequency signals. However, these priors are weak in the sense that they do not explicitly exploit the disparity pattern of the light field signal. To explicitly capture this structure, Levin et al. [10] model the light field data using a 1-D depth variable and a 2-D texture variables. They proposed a GMM model for the light field data, in which conditioned on the depth variable, the texture variable is a 2-D Gaussian random variable. During reconstruction, the depth variable is first estimated using a MRF formulation and then the texture estimation becomes a simple linear problem. Further modifications to this prior has been proposed in [9, 11]. Our patch prior model is very similar to the prior proposed in [10]. However, the prior in [10] being a global prior, the inference is computationally expensive. On the other hand, our prior being defined on small patches, both learning and inference are computationally efficient. Also, our inference step can be parallelized by inferring each patch independently.

Among light field processing tasks, (spatial) super-resolution has been considered in [14, 5, 3, 2]. In [14], it is first detected whether the sub-images under each microlens of the plenoptic camera are flipped or not, and then their central part is scaled up assuming that the scene is an equifocal plane at a user-defined depth. In [3], the image formation model of the plenoptic camera is derived and the light field data is modeled as an autoregressive (AR) model. The super-resolved data is then obtained using Bayesian inference technique. A multi-view denoising algorithm has been proposed in [21], where similar (noisy) patches from different viewpoints are grouped together using depth estimates of the patches and are denoised together by using PCA and tensor analysis. Recently, Yu et al. [20] have shown that dynamic depth of scene effects can be produced from a stereo camera. They recover a high resolution depth map from the stereo images and use this depth map to synthesize the light field of the scene.

Patch priors are quite popular in the image processing [4, 15, 7, 19, 22]. The main advantage of using these priors is that learning and inference are very efficient as compared to the global image priors. Further, GMM patch priors have been shown to be very effective in solving many image processing problems [19, 22].

3. Problem Definition

We provide a common framework for all light field processing tasks whose observation models are linear. Let x represent the light field data that we want to reconstruct and let the observation model of the processing task be given by

$$y = Hx + n, \tag{1}$$

where y is the observed data and n is the observation noise which we will assume to be Gaussian distributed. As an example, for the denoising problem H is the identity matrix and y is the noisy light field data. Our goal is to estimate the original light field data from the observation y. Since, we take a patch based approach, y denotes extracted patches from the observed data and x represents the corresponding light field patch that we want to estimate. To obtain the complete light field data, we estimate the individual patches separately and then aggregate them later. In section 4, we study the structure of light field patches, which naturally suggests a GMM prior for patches. In section 5, we describe our patch prior and its associated reconstruction algorithm.

4. Low-dimensional Structure of Light Field Patches

In this section, we first describe the disparity pattern in light field data. We then use these patterns to compute the intrinsic dimension of light field patches corresponding to different scene depths. This computations provide us with

an estimate of the minimum number of observations needed to reliably estimate light field patches of different disparity values.

Regular light field data (such as those captured by the plenoptic camera [16]) can be considered as a 2-D array of images, with the two dimensions representing the angular dimensions u and v. Figure 1 shows various slices (X-Y, X-U and Y-V) of a light field data. The specific intensity patterns in the X-U and Y-V slices indicate that many of the pixels have the same values. These pattern arise because a diffuse scene point, when observed from different angular locations (views) in the U-V plane, gets mapped to slightly different spatial locations in the X-Y plane. For example, an in-focus scene point (a scene point which is in the plane of focus of the main lens) gets mapped to the same spatial location in all the images. Whereas, an out-of-focus scene point gets mapped to slightly different spatial locations. Further, the amount of spatial shifts (pixel disparities) between any two consecutive images in the array is the same. The amount of disparity is an indication of the depth of the scene from the focal plane.

As discussed earlier, we process the light field data by extracting many small patches from it, lets say of size $n_1 \times n_2 \times s_1 \times s_2$. For small patch sizes, we can assume that the whole patch can be characterized by a single disparity value. The disparity value of a patch characterizes its intrinsic dimension (defined as the minimum number of variables to represent the patch). Though the ambient dimension of all the patches are the same, the intrinsic dimensions are quite different. For example, consider a patch with disparity value 0: the $X-Y$ slices (there are $s_1 s_2$ such slices) of this patch are all the same and hence the intrinsic dimension is only $n_1 n_2$ (whereas the ambient dimension is $n_1 n_2 s_1 s_2$). Figure 2 shows how to compute the intrinsic dimension of a 2-D light field patch. It is easy to extend this argument to the 4-D light field case and verify that the intrinsic dimension of a patch with disparity value of d pixels (between consecutive images in the 2-D image array) is given by

$$(n_1 + d(s_1 - 1))(n_2 + d(s_2 - 1)). \qquad (2)$$

The intrinsic dimension provides us with a lower bound on the number of observations needed to reconstruct the light field reliably. Figure 3 shows the ratio of intrinsic to ambient dimensions for light field patches of size $16 \times 16 \times 5 \times 5$ for different pixel disparity values. We have plotted two curves, one computed from Equation (2), and the other computed empirically from the rank of the patch covariance matrices (for each disparity value, we learn a patch covariance matrix). Note that the two curves agree with each other exactly. From this figure, we can conclude that if the scene has a maximum disparity of $+/-5$ (and we use $16 \times 16 \times 5 \times 5$ patches), then we can reliable recover the light-field data even if the observation model is underdetermined by a factor of 5 (which corresponds to the ratio of intrinsic to ambient dimension being equal to 0.2).

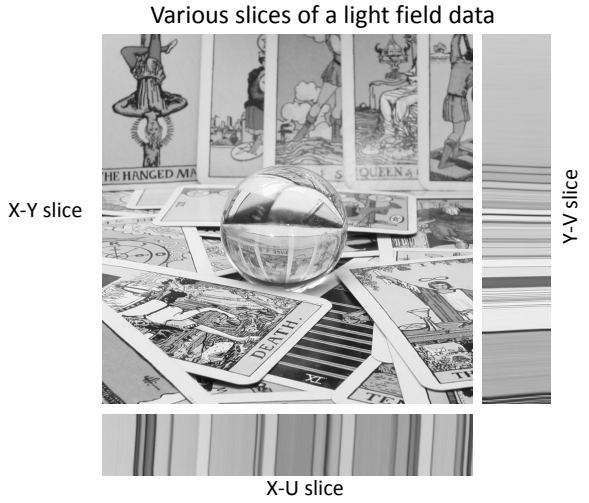

Figure 1: Various slices (X-Y, X-U and Y-V) of a light field data. The specific intensity patterns in the X-U and Y-V slices indicate that many of the pixels have the same values. Hence, the intrinsic dimension (defined as the minimum number of variables to represent a signal) of the light field patches are much lower than its ambient dimension.

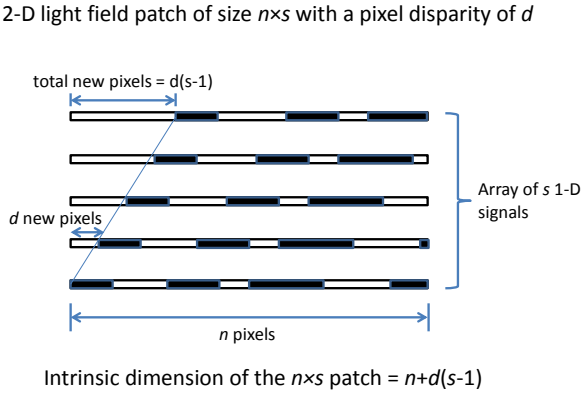

Figure 2: This figure shows how to compute the intrinsic dimension of a 2-D light field patch. It is easy to extend this argument to the 4-D light field case and verify that the intrinsic dimension of a patch of size $n_1 \times n_2 \times s_1 \times s_2$ and disparity value d is given by Equation 2.

5. GMM Patch Prior

We present our GMM patch prior model and a fast reconstruction algorithm based on the MAP criterion. As

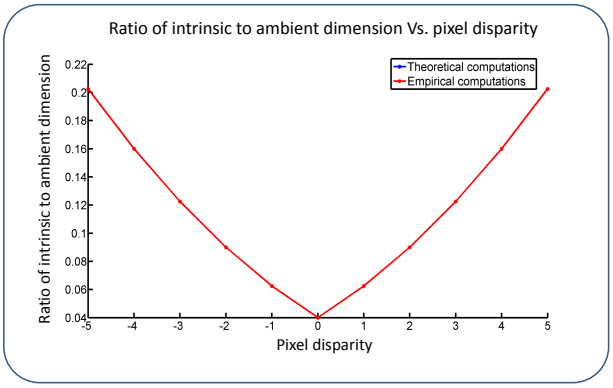

Figure 3: Ratio of intrinsic to ambient dimensions of $16 \times 16 \times 5$ light field patches for different pixel disparity values: We show two plots, one computed from Equation (2), and the other computed empirically from the rank of the patch covariance matrices (for each disparity value, we learn a patch covariance matrix). Note that the two curves agrees with each other exactly. From this plot, we can also conclude that if the scene has a maximum disparity of $+/-5$ (and we use $16 \times 16 \times 5$ patches), then we can reliable recover the light field data even if the observation model is under-determined by a factor of 5 (which corresponds to the ratio of intrinsic to ambient dimension being equal to 0.2).

discussed in the previous section, the light field patches of different disparities have different intrinsic dimensions and hence they are fundamentally different. Keeping this difference in mind, we learn a Gaussian prior for each of the (m quantized) disparity values $d_i, i = 1, 2, ..., m$. That is,

$$P(x|d_i) = \mathcal{N}(x|m_{d_i}, \Sigma_{d_i}), \tag{3}$$

where m_{d_i} and Σ_{d_i} are the mean vector and covariance matrix respectively for patches of disparity value d_i. We assume each of the disparities to be equally likely and thus the overall patch prior $P(x)$ is a GMM:

$$P(x) = \frac{\sum_{i=1}^{m} P(x|d_i)}{m}. \tag{4}$$

We learn the mean vector and covariance matrix for each of the disparity values by generating artificial light field data. We obtain the artificial light field data by generating a 2-D array of images by shifting it by specified disparity values.

As discussed in section 3, any (linear) processing task can be represented by Equation (1). If we assume the observation noise to be i.i.d Gaussian distributed, then the likelihood distribution $P(y|x)$ is given by

$$P(y|x) = \mathcal{N}(y|Hx, \sigma^2\mathcal{I}). \tag{5}$$

Given an observation patch y, one can use the maximum aposterior criterion (MAP) to estimate x. However, computing the MAP estimator of for the GMM prior is computationally expensive. Thus, we instead propose an fast (but approximate) algorithm, where we first estimate the disparity value using a subspace projection operation and then use the MAP criterion to estimate the light field patch x for the estimated disparity value.

For each disparity value d_i, we compute the top PCA subspace, i.e., the subspace spanned by the top few eigenvectors of the covariance matrix Σ_{d_i}. Lets denote the subspace corresponding to disparity d_i by Q_i. Under the transformation operation H in Equation (1), the subspace will be transformed to HQ_i. Given an observation patch y, we project it to all the transformed subspaces HP_i and compute the projection residual r_i

$$r_i = \min_{\alpha} ||y - HQ_i\alpha||_2. \tag{6}$$

The disparity value corresponding to the minimum projection residual r_i is taken to be the estimated disparity value $\hat{d}$. Using this disparity value $\hat{d}$, we then compute the MAP estimator of x. The MAP estimator for Gaussian prior $\mathcal{N}(x|m_{\hat{d}}, \Sigma_{\hat{d}})$ and Gaussian likelihood Equation (5) is given by:

$$\hat{x} = m_{\hat{d}} + \Sigma_{\hat{d}}H^T(H\Sigma_{\hat{d}}H^T + \sigma^2\mathcal{I})^{-1}(y - Hm_{\hat{d}}). \tag{7}$$

This is also known as the linear minimum mean square estimator (LMMSE) of x. Figure 4 summarizes our algorithm.

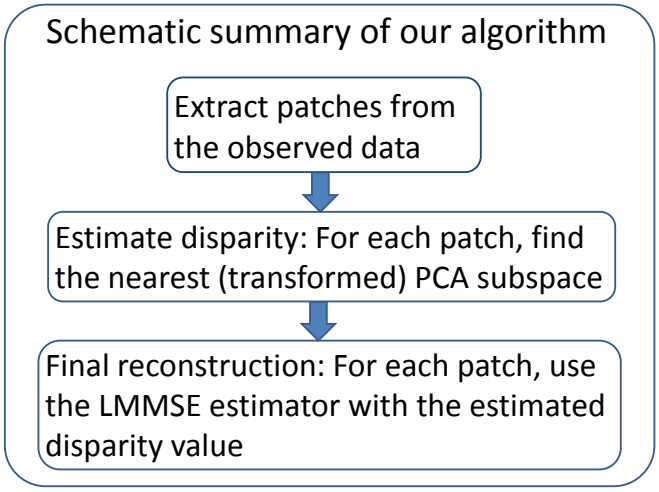

Figure 4: Summary of our algorithm: Given the observed data, we first extract patches from it. We then estimate the disparity value for each patch by projecting the patch onto the low-dimensional subspaces corresponding to different disparity values and finding the one that is closest. Once we have estimated the disparity value for a patch, we then reconstruct it using the LMMSE estimator corresponding to that particular disparity value.

6. Applications of the GMM Patch Prior

We apply our light field GMM patch prior to denoise light field data, create many refocused images from a

stereo image pair and (spatially) super-resolve the light field data. We learn GMM prior for light field patches of size $16 \times 16 \times 5 \times 5$ and disparity values ranging from -5 to 5 with a quantization of 1. While estimating the disparity value of patches, we use the top 50 eigen-vectors, i.e., $rank(Q_i) = 50$. We extract overlapping patches from the observed data and aggregate the reconstructed light field patches by simple averaging operation.

6.1. Light field Denoising

Denoising is an important processing task in the context of light field data as some of the light field camera architectures the light throughput is low. To simulate large amount of noise, we add noise to the light field data (taken from the Stanford light field archive). The light field data we use for our experiments has angular dimension of 5×5. We extract (overlapping) patches of size $16 \times 16 \times 5 \times 5$ from the noisy light field data, estimate the disparity value for each of the patches by using the subspace projection method and then denoise them using the LMMSE algorithm. Finally, we aggregate all the patches to get our final denoised light field data. Figure 5 shows the X-Y, X-U and Y-V slices of the original light field data, noisy data and the denoised data. It also shows the estimated disparity map. Note that our denoising approach recovers the correct disparity structure of the data.

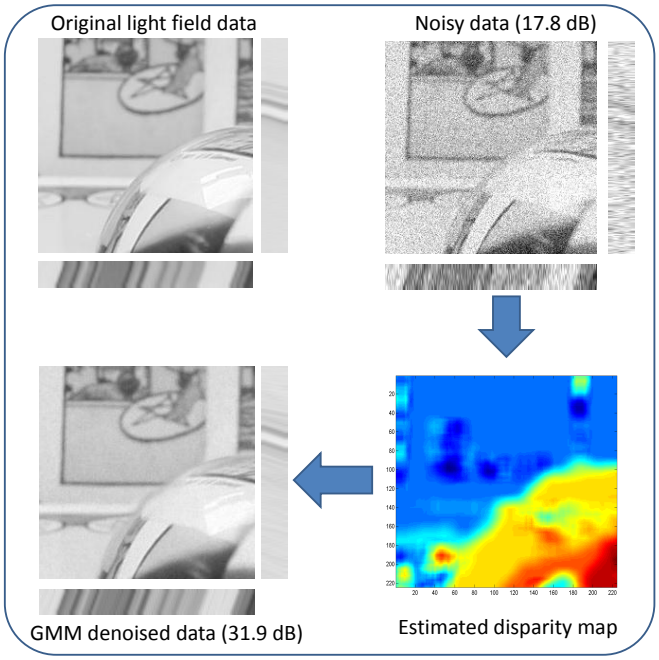

Figure 5: Light field denoising using our GMM patch prior model: We extract patches from the noisy input data, estimate the disparity value for each of the patches using our subspace projection method and then denoise them using the LMMSE algorithm. The figure shows X-Y, X-U and Y-V slices of the original light field data, noisy data and the denoised data. We also show the estimated disparity map. Note that our denoising approach recovers the corect disparity pattern of the data.

6.2. Refocusing from Stereo Pair

One of the main applications of capturing light field is to obtain refocused images at different depths after capturing a scene. Here we show that by capturing just two images (for example from a stereo camera), we can perform refocusing at many depth values. In our experiments, we use two images from a light field data of angular resolution 5×5, see Figure 6. Given the two input images, we reconstruct the light field data and then perform refocusing. We extract 16×16 patches from each of the images, estimate the depth map and then reconstruct the light field patches. From the reconstructed light field data, we then obtain the refocused images at many depth values, see figure 6.

6.3. Light Field (Spatial) Super-resolution

Since light field cameras trades off the spatial resolution for angular resolution, there is a need to develop good super-resolution algorithms. In this experiment we demonstrate that we can spatially super-resolve the light field data by a factor of 4. From the input low-resolution light field data, we extract patches of size $4 \times 4 \times 5 \times 5$ and using our GMM patch framework reconstruct the corresponding $16 \times 16 \times 5 \times 5$ patches. Aggregating these patches give us the super-resolved light field data. Figure 7 shows the super-resolved data obtained by our approach and by the standard bicubic interpolation (we perform bicubic interpolation on each of the low-resolution images/views of the light field data).

7. Conclusions

We provide an unified framework for solving many light field processing tasks. We use a patch based approach where we model the light field patches using a GMM prior. First, we analyze light field patches of diffuse scenes and show that patches with the same disparity value lie on a low-dimensional subspace. We also compute the dimensionality of each such disparity subspaces and show that it varies quadratically with the disparity value. This analysis gives us an indication of the amount of under-determinacy we can tolerate in a processing task. We then model the patches as Gaussian random variables conditioned on its disparity value, thus, effectively leading to a GMM model. Given a processing task (whose observation is linear), we provide a two step algorithm for estimating the original light field patches from the observed patches. In the first stage, we estimate the disparity value of the observed patches using a fast subspace projection algorithm and in the second stage we use the LMMSE algorithm to finally estimate the light

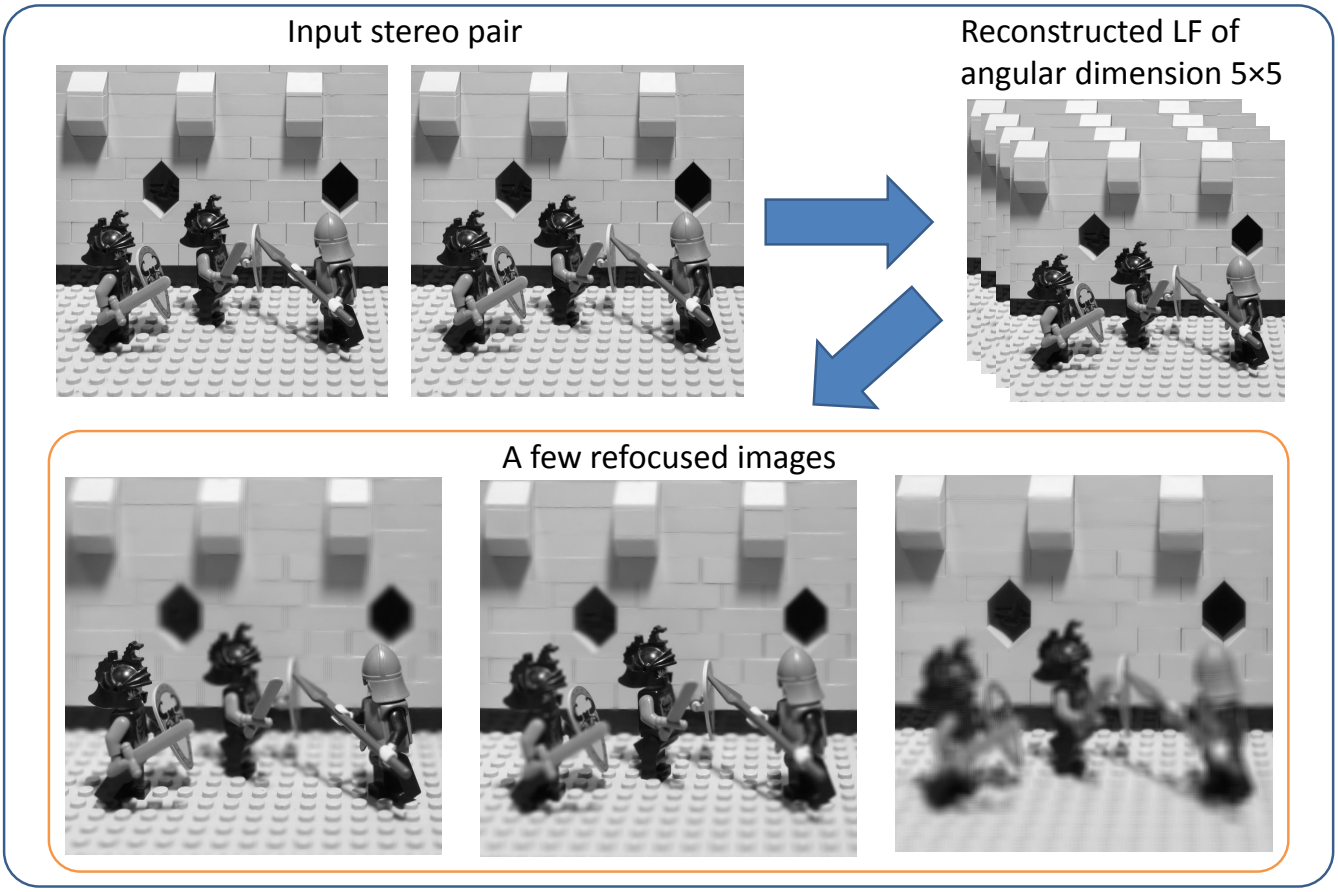

Figure 6: Our goal is to perform post-capture refocusing at many depth values from just two images (stereo pair). Using our GMM patch prior, we reconstructed the light field data of angular resolution 5×5 and then obtain refocused images at many depth values. Here we show three of the refocused images.

field patches. We also demonstrated the effectiveness of our approach in solving the light field denoising problem, and the light field angular and spatial super-resolution problems.

Currently, the proposed algorithm assumes a diffuse scene and that the patches are small enough so that the effect of depth discontinuities are negligible. In future, we would like to explore more sophisticated priors that can handle complex reflectance models and depth discontinuities.

Acknowledgements

This work was supported by NSF Grants NSF-IIS: 1116718, NSF-CCF:1117939 and by a gift from Samsung Telecommunications America.

References

[1] E. H. Adelson and J. Y. A. Wang. Single lens stereo with a plenoptic camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2), 1992. 2

[2] 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*, 34, 2012. 2

[3] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *IEEE International Conference on Computational Photography*, 2009. 2

[4] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15, Dec. 2006. 2

[5] T. Georgiev, G. Chunev, and A. Lumsdaine. Superresolution with the focused plenoptic camera. *SPIE Electronic Imaging*, Jan 2011. 2

[6] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *In Proceedings of SIGGRAPH*, pages 43–54. ACM, 1996. 2

[7] V. K. K. Dabov, A. Foi and K. Egiazarian. Image restoration by sparse 3d transform-domain collaborative ltering. *SPIE Electronic Imaging*, 2008. 2

[8] S. M. Kay. Fundamentals of statistical signal processing: Estimation theory. *Prentice-Hall, USA*, 1993. 2

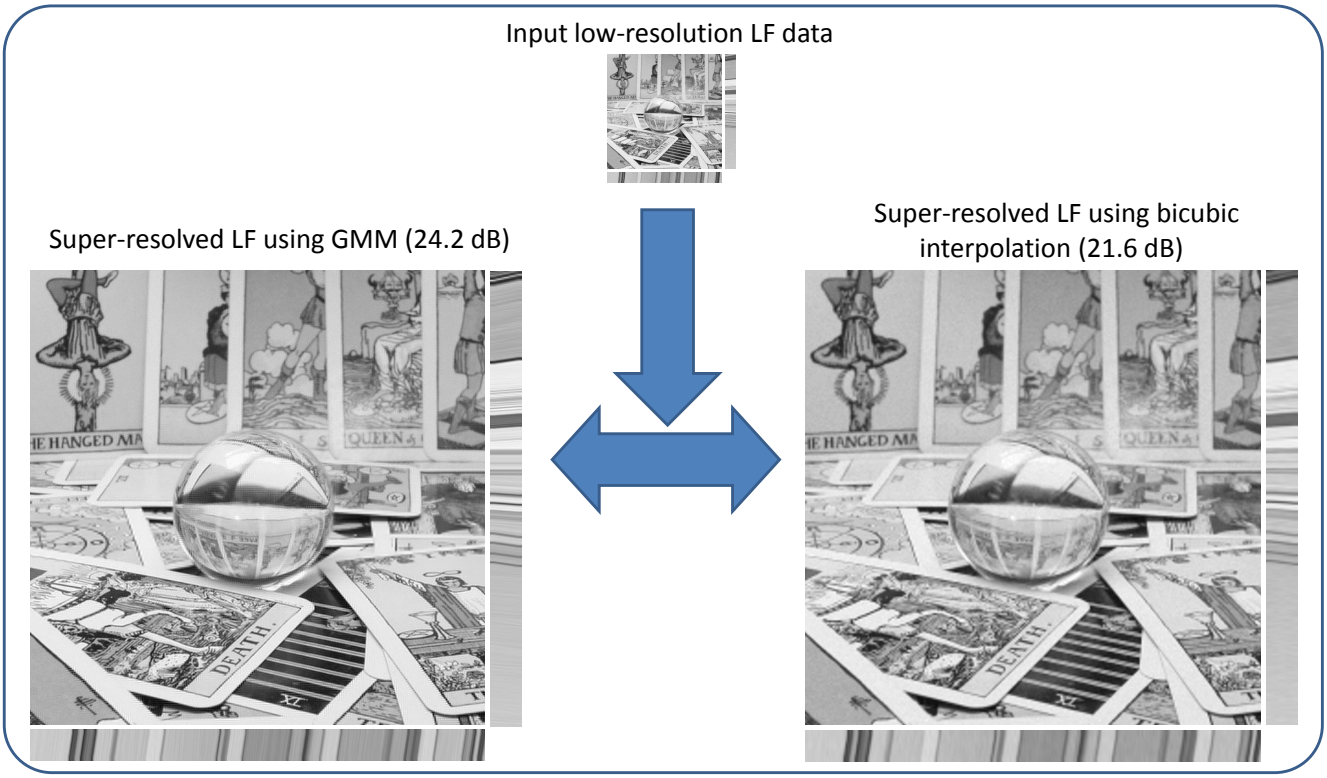

Figure 7: Light field cameras have low spatial resolution and hence we need good super-resolution algorithms. Here, we show that using our framework we can super-resolve the input data by a factor of four. We also compare our algorithm with the standard bicubic interpolation method (we perform bicubic interpolation on each of the low-resolution images/views of the light field data).

[9] A. Levin and F. Durand. Linear view synthesis using a dimensionality gap light field prior. In *CVPR*, 2010. 2

[10] A. Levin, W. T. Freeman, and F. Durand. Understanding camera trade-offs through a bayesian analysis of light field projections. In *ECCV*, 2008. 2

[11] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman. 4d frequency analysis of computational cameras for depth of field extension. *ACM Trans. Graph.*, 28(3), 2009. 2

[12] M. Levoy and P. Hanrahan. Light feld rendering. *SIGGRAPH*, 1996. 1, 2

[13] C. Liang, T. Lin, B. Wong, C. Liu, and H. Chen. Programmable aperture photography: multiplexed light field acquisition. In *ACM SIGGRAPH 2008 papers*, pages 1–10. ACM, 2008. 1

[14] A. Lumsdaine and T. Georgiev. Full resolution lightfield rendering. *Adobe Tech Report*, January 2008. 2

[15] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *Computer Vision, 2009 IEEE 12th International Conference on*, 2009. 2

[16] R. Ng, M. Levoy, G. Duval, P. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. April 2005. 1, 2, 3

[17] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transactions on Graphics*, 26(3), 2007. 1

[18] 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.*, 24(3), jul 2005. 1, 2

[19] G. Yu, G. Sapiro, and S. Mallat. Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity. *arXiv:1006.3056*, 2010. 2

[20] Z. Yu, C. Thorpe, X. Yu, S. Grauer-Gray, F. Li, and J. Yu. Dynamic depth of field on live video streams: A stereo solution. *In proceedings of Computer Graphics International (CGI)*, 2011. 2

[21] L. Zhang, S. Vaddadi, H. Jin, and S. Nayar. Multiple view image denoising. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2009. 2

[22] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011. 2