Light Field Photography with Lytro Imaging: A Power Tool for an Image Revolution

Komal
Assistant Professor, Rama University
k.komal208@gmail.com

Abstract: Since 1996, research on light fields has followed a number of lines. On the theoretical side, researchers have developed spatial and frequency domain analyses of light field sampling and have proposed several new parameterizations of the light field, including surface light fields and unstructured Lumigraphs. In this paper, we present a novel descriptor in light field to tackle that how to avoid the invading of the attack in the biometric system, such as 2D printed photos. The proposed light field histogram of gradient (LFHoG) descriptor is derived from three directions, including vertical, horizontal and depth. Different with traditional HoG in 2D image, the gradient in depth direction is distinctive in light field.

Keywords :Lytro camera, discriminative descriptor, live face detection, light field image, light field histogram of gradient descriptor, light field histogram of gradient descriptor.

I. INTRODUCTION

The light field imaging techniques [1] have been developed quickly in recent years and commercial light field cameras, such as Lytro and Raytrix have been available in the market. The reason why light field imaging is popular is that it can collect more angular information for an incident ray while traditional imaging only records one angular information. As a result, the depth estimation becomes very easy and it makes many applications possible, e.g. image super-resolution [2], refocusing [3] and 3D reconstruction [4]. The light field can also be utilized in the biometric system [5] to enhance face/iris recognition by its abundant information and it is becoming a developing tendency in the biometric system to equip light field cameras. In place of film and pinholes, the Lytro camera uses a thin sheet containing thousands of micro lenses, which are positioned between a main zoom lens and a standard 11-megapixel digital image sensor. The main lens focuses the subject onto the sheet of micro lenses. Each micro lens in turn focuses on the main lens, which from the perspective of a micro lens is at optical infinity.

Figure (1) First Image capture from the normal lens 2.7 megapixel digital camera and second image taken from 11 megapixel Lytro camera lens.

With a glass array of around 20 000 micro lenses in its €20 000 (about $26 500) R11 camera, for example, Raytrix manages to produce 2.7-megapixel still images from an 11-megapixel sensor and video at up to six frames per second. Unlike with Lytro, though, the number-crunching hardware needed for Plenoptic 2.0 cannot fit into a stylish anodized case. A Raytrix camera must be linked through a gigabit Ethernet cable to a PC that contains a high-end Nvidia GeForce GTX 580 graphics card, which itself costs more than the entire Lytro camera.

II. RELATED WORK

The problem of the light field's limited resolution has been extensively studied in the past and several powerful methods for increasing the resolution in both angular [Levin and Durand 2010; Shi et al. 2014;Wanner and Goldluecke 2014] and spatial [Bishop et al. 2009;Cho et al. 2013] domains have been proposed. For brevity, we only focus on the approaches that are designed for angular super-resolution. We start by reviewing the algorithms that specifically work for light fields and then explain the approaches that perform view synthesis for general scenes and objects.

Light Field Super-resolution – Levin and Durand [2010] use a prior based on the dimensionality gap to reconstruct the full 4D light field from a 3D focal stack sequence. Shi et al. [2014] leverage scarcity in the continuous Fourier spectrum to reconstruct a dense light field from a 1D set of viewpoints. Schedl et al. [2015] reconstruct a full light field using multidimensional patches from a sparse set of input views. These methods require the input samples to be captured with a specific pattern and are not able to synthesize novel views at arbitrary positions. Marwah et al. [2013] propose a dictionary-based approach to reconstruct light fields from a coded 2D projection. However, their method requires the light fields to be captured in a compressive way. Mitra and Veeraraghavan [2012] introduce a patch-based approach where they model the light field patches using a Gaussian mixture model. However, this method is not robust against noise, and struggles on low-quality images taken with commercial light field cam- eras. Zhang et al. [2015] propose a phase-based approach to re- construct light fields. However, their method is limited since it is designed for a micro-baseline stereo pair. Moreover, their approach is iterative, which is often slow and prevents its usage in practice. Yoon et al. [2015] perform spatial and angular super-resolution on light fields using convolution neural networks (CNN). However, their method can only increase the resolution by a factor of two, and is not able to synthesize views at arbitrary locations. Layered patch-based synthesis has been proposed by Zhang et al. [2016] for various light field editing applications. Although they show impressive results for applications like hole-filling and reshufflingRecently, Wanner and Goldluecke [2014] proposed an optimization approach to reconstruct images at novel views from an input light field. Given the depth estimates at the input views, they reconstruct novel views by minimizing an objective function which maximizes the quality of the final results. Although their method produces reasonable results on dense light fields, for sparse input views. We believe this is because of two main reasons. First, they estimate the disparity at the input views as a pre process, independently of the view synthesis process. However, even state- of-the-art light field disparity estimation techniques [Wang et al. 2015;Jeon et al. 2015] are not typically designed to maximize the quality of synthesized views, and thus, they are not suitable for this application. Second, Wanner and Goldluecke's method assumes that the images are captured under ideal conditions.

View Synthesis for Scenes – View synthesis has a long history in both vision and graphics. One category of approaches [Eisemann et al. 2008;Goesele et al. 2010;Chaurasia et al.2011;Chaurasia et al. 2013] synthesizes novel views of a scene in a

two-step pro- cess. These methods first estimate the depth at the input views and use the depth to warp the input images to the novel view. They then produce the final image by combining these warped images. These approaches typically use multi-view stereo algorithms (e.g., PMVS by Furukawa et al. [2010]) to estimate depth and are not suitable for light fields with a narrow baseline. In our system, we also have depth and color estimation components. However, unlike these ap- proaches, we use machine learning to model these two components. Furthermore, inspired by Fitzgibbon et al.'s approach [2003], we train both our disparity and color estimation models by directly minimizing the appearance error.

Deep Stereo – Flynn et al. [2016] has recently proposed a deep learning method to perform view synthesis on a sequence of images with wide baselines. They first project the input images on multi- ple depth planes. They then estimate the pixel color and weight of the image at each depth plane from these projected images. Finally, they compute a weighted average of the estimated pixel colors to obtain the final pixel color. Comparing to this approach, our sys- tem has several key differences. First, our method is specifically designed for light fields, which have much narrower baselines and more regular camera positions. Second, unlike their approach, our system explicitly estimates the disparity which could potentially be used in other applications. Finally, our system is significantly faster than their method.

III. PROPOSED LEARNING-BASED ALGORITHM

Given a sparse set of input views $L_{p1}, \cdots, L_{pN}$ and the position of the novel view q, our goal is to estimate the image at the novel view L_q. Formally, we can write this as:

$$L_q = f(L_{p1}, \cdots L_{pN}, q), \quad (1)$$

where pi and q refer to the (u, v) coordinates of the input and novel view, respectively. Here, f is a function which defines the rela- tionship between the input views and the novel view. This relation- ship is typically very complex as it requires finding connections between all the input views, and collecting appropriate information from each image based on the position of the novel view. Inaccu- racies such as noise and optical distortions in consumer light field cameras further add to the complexity of this relationship. Therefore, we propose to learn this relationship. Inspired by the recent success of deep learning in a variety of applications, we propose to use convolutional neural networks (CNN) as our learning model. A straightforward way to doso is to directly model the function f with a CNN. In this case, the CNN takes the input views as well as the position of the novel view and outputs the image at the novel view. This is mainly due to the fact that the relationship is complex and requires the network to find connections between distant pixels, which makes the training difficulties vs. seconds). This shows the efficiency of our system, validating more practical usage.

We make the training more tractable by following the pipeline of existing view synthesis techniques [Chaurasia et al. 2011;Chaura-sia et al. 2013] and breaking the system down into disparity and color estimation components. Our main contribution is to use machine learning to model each component and train both models simultaneously by minimizing the error between the synthesized and ground truth images. In our system, we first estimate the disparity at the novel view from a set of features extracted from the sparse set of input views:

$$D_q = g_d(K), \quad (2)$$

where D_q is the estimated disparity at the novel view, K represents a set of features including the mean and standard deviation of warped images at different disparity levels. Moreover, g_d defines the relationship between the input features and the disparity

which we model using a CNN. The estimated disparity is then used to warp the input images to the novel view. Specifically, we perform a backward warp by sampling the input images based on the disparity at the novel view (see Eq.3). Finally, we estimate the image at the novel view using a set of input features including all the warped images, the estimated disparity, and the position of the novel view:

$$Lq = gc\ (H), \qquad (3)$$

where H represents our feature set and gc defines the relationship between these features and the final image. The overview of our system is shown in Fig.3. In the next sections we describe the disparity estimator (Eq.2) and the color predictor (Eq.3) in detail.

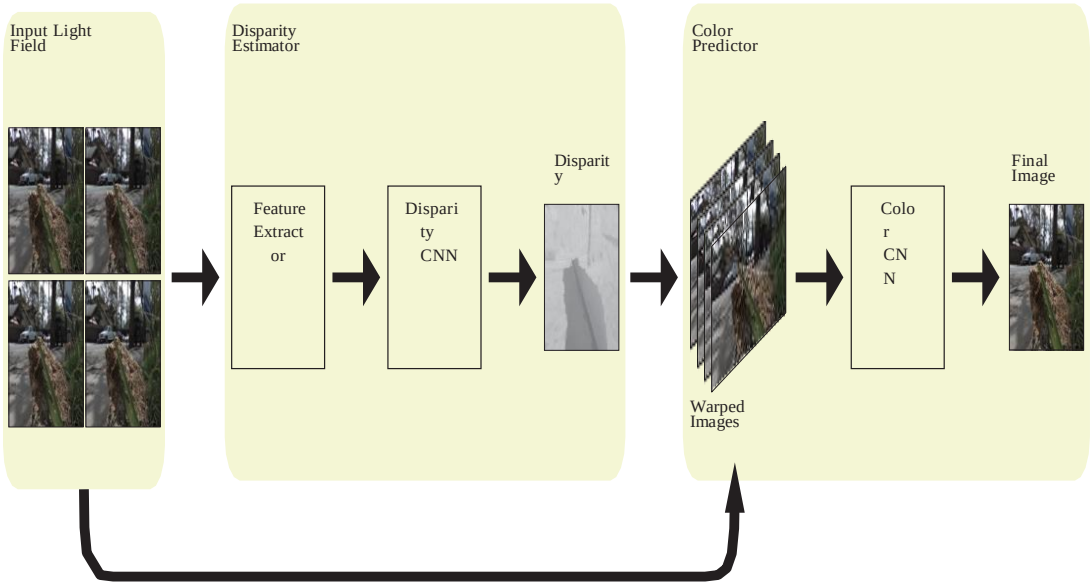

Figure 2 : Our system consists of disparity estimator and color predictor components which we model using two sequential CNNs. In our system, we first extract a set of features from the sparse input views. We then use the first CNN to estimate the disparity at the novel view. We then use this disparity to warp (backward) all the input views to the novel view. Our second CNN uses all the warped images along with a few other features to generate the final image

IV. CONCLUSIONS AND FUTURE WORK

We have presented a novel learning-based approach for synthesizing novel views from a sparse set of input views captured with a consumer light field camera.

We show the result of our approach on a variety of scenes using only the four corner sub-aperture images captured with a Lytro camera. In the future, we would like to investigate the possibility of using our system for generating high dynamic range light fields from a set of views with different exposures. Moreover, it would be interesting to extend our system to work with any number of input views

ACKNOLEDGEMENT

I would like to thanks to our team supporter Mr. Ashutosh and my family for their kind supports**.**

REFERENCES

[1] ADELSON, E. H., AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. *IEEE PAMI 14*, 2, 99–106.

[2] BISHOP, T. E., ZANETTI, S., AND FAVARO, P. 2009. Light field super resolution. In *IEEE ICCP*, 1–9.

[3] BURGER, H. C., SCHULER, C. J., AND HARMELING, S. 2012. Image demising: Can plain neural networks compete with BM3D? In *IEEE CVPR*, 2392–2399.

[4] CHAURASIA, G., SORKINE, O., AND DRETTAKIS, G. 2011. Silhouette-aware warping for image-based rendering. In *EGSR*, 1223–1232.

[5] CHAURASIA, G., DUCHENE, S., SORKINE-HORNUNG, O., AND DRETTAKIS, G. 2013. Depth synthesis and local warps for plausible image-based navigation. *ACM TOG 32*, 3, 30:1–30:12.

[6] CHO, D., LEE, M., KIM, S., AND TAI, Y.-W. 2013. Modeling the calibration pipeline of the lytro camera for high quality light- field image reconstruction. In *IEEE ICCV*, 3280–3287.

[7] DONG, C., LOY, C. C., HE, K., AND TANG, X. 2014. Learning a deep convolution network for image super-resolution. In *ECCV*, 184–199.

[8] DOSOVITSKIY, A., SPRINGENBERG, J. T., AND BROX, T. 2015. Learning to generate chairs with convolutional neural networks. In *IEEE CVPR*, 1538–1546.

[9] EISEMANN, M., DE DECKER, B., MAGNOR, M., BEKAERT, P., DE AGUIAR, E., AHMED, N., THEOBALT, C., AND SELLENT, A. 2008. Floating textures. *CGF 27*, 2, 409–418. FITZGIBBON, A., WEXLER, Y., AND ZISSERMAN, A. 2003. Image-based rendering using image-based priors. In *IEEE ICCV*, 1176–1183 vol.2.

[10] FLYNN, J., NEULANDER, I., PHILBIN, J., AND SNAVELY, N 2016. Deepstereo: Learning to predict new views from the worlds imagery. In *IEEE CVPR*, 5515–5524.

[11] FURUKAWA, Y., AND PONCE, J. 2010. Accurate, dense, and ro- bust multiview stereopsis. *IEEE PAMI 32*, 8, 1362–1376.

[12] GEORGIEV, T., ZHENG, K. C., CURLESS, B., SALESIN, D., NAYAR, S., AND INTWALA, C. 2006. Spatio-angular resolution tradeoffs in integral photography. In *EGSR*, 263–272.