Ding Luo*, Thomas Längle, and Jürgen Beyerer

Compressive shape from focus based on a linear measurement model

Komprimierte Tiefenmessung mit variablem Fokus auf Basis eines linearen Messmodells

DOI 10.1515/teme-2017-0007
Received January 30, 2017; revised March 21, 2017; accepted April 6, 2017

Abstract: Estimation accuracy of conventional shape from focus techniques is strongly coupled with the number of images in the focal stack, limiting the measurement speed. In this article, a novel compressive shape from focus scheme is proposed with an exemplary algorithm based on modified Laplacian operator and principal component analysis. Simulation with synthetic focal stacks have demonstrated comparable results to the conventional method. A test with 6 compressively captured images achieves the same level of performance to that of the conventional method with 100 images. Several other focus measure algorithms are also implemented and tested under the compressive scheme, which demonstrates the wide applicability of the proposed method.

Keywords: Shape from focus, modified Laplacian, PCA.

Zusammenfassung: Die Schätzgenauigkeit der herkömmlichen Tiefenmessung mit variablem Fokus ist mit der Anzahl der Bilder im Fokusstapel gekoppelt, und das begrenzt die Messgeschwindigkeit. In diesem Artikel wird eine neuartige komprimierende Methode mit einem beispielhaften Algorithmus basierend auf einem modifizierten Laplace-Operator und der Hauptkomponentenanalyse vorgeschlagen. Die Simulation mit synthetischen Fokusstapeln hat Ergebnisse gezeigt, die mit der herkömmlichen Methode vergleichbar sind. Der Test mit 6 komprimiert erfassten Bildern erreicht dieselbe Leistung wie die herkömmliche Methode mit 100 Bildern. Mehrere andere Fokusmessalgorithmen werden auch unter dem Kompressionsschema implementiert und getestet, was die breite Anwendbarkeit des vorgeschlagenen Verfahrens demonstriert.

Schlüsselwörter: Tiefenmessung mit variablem Fokus, modifizierter Laplacian-Operator, Hauptkomponentenanalyse.

***Corresponding author: Ding Luo,** Karlsruher Institut für Technologie (KIT), Lehrstuhl für Interaktive Echtzeitsysteme, Karlsruhe, Germany, e-mail: ding.luo@iosb.fraunhofer.de
Thomas Längle, Jürgen Beyerer: Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB), Karlsruhe, Germany

1 Introduction

Depth estimation based on an imaging system has been a widely studied topic in the area of computer vision and image processing. Generally, existing methods can be classified into active methods and passive methods. Active methods involve the projection of an optical probe onto the target scene, often in the form of laser beam or illumination pattern [9]. The 3D profile of the target scene is reconstructed with the information in the scattering/reflection of the optical probe captured by the imaging system. The requirement of the additional projection/illumination system will increase the complexity and cost of the active methods, inevitably limiting their applicability. In situations where physical interaction with the scene is not allowed, passive methods are applied by taking images of the scene without additional illumination. Various depth cues in the captured images have been proposed by researchers, including stereopsis [4], shading [13], focus [6], etc., which are used to reconstruct the 3D information. In this paper, the usage of focus as a cue for depth measurement will be studied and discussed.

Research focuses in this area are mainly placed upon two topics, the design of robust focus measure operators and the development of estimation algorithms. Pertuz et al. [7] made an extensive survey and comparison of popular focus measure operators for shape from focus. Apart from the operators listed in the above survey, more complex operators are being developed constantly not only for shape from focus but also for sharpness estimation as a more general topic, such as the S_3 operator by Vu et al. [12], which utilizes both spatial and spectral information in color images. Conventional estimation algorithms involve finding the maximum focus position from the focal stack for each pixel. A widely accepted method is to take a Gaussian model as proposed by Nayar et al. [7]. Alternatively, other fitting methods have also been studied, such as quadratic and polynomial fits [10]. With the recent development of machine learning and optimization algorithms, more sophisticated methods have been proposed by breaking the isoplanatic restriction [10], such as surface fitting and optimization by neural networks [1], and total variation regularization [3]. It can be seen from the listed literature that the design of focus measure and the development of estimation algorithm are often conducted simultaneously in a holistic manner in order to improve the performance of the overall method.

Unlike shape from defocus techniques where the blur kernel is assumed known, shape from focus techniques generally require a minimum number of image samples along the focal axis in order to perform robust estimation, which are achieved by either shifting the focal plane or changing the relative distance between the camera and the scene. When large numbers of images are required, such shift/movement commonly leads to slow measurement speed and bulky systems. Additionally, the large number of the images, which are needed for evaluation, adds to the data transfer and computational cost. To tackle this problem, a novel shape from focus scheme is proposed and discussed. Although only the algorithm is described in detail and simulated in this paper, potential hardware implementations will be introduced briefly in the last section.

In Section 2, theoretical background is given with a focus on the linear measurement model, which serves as the core of this new shape from focus (SFF) scheme. Section 3 describes the proposed algorithm with an example and the results from simulation tests are presented in Section 4.

2 Theoretical background

2.1 Shape from focus

The key concept of recovering depth information from a focal stack is the relationship between focused and defocused images. In a thin lens model, image points that are sharply projected on an image plane fulfill the Gaussian lens equation:

$$\frac{1}{o} + \frac{1}{i} = \frac{1}{f} \tag{1}$$

where o is the distance of the object point from the lens plane, i denotes the distance of the focused image from the lens plane and f represents the focal length of the optical system. When a detector is placed at the focus and the object is moved away from the original position, the image of the object will be blurred. The degree of blurring depends on how far away the object is from the in-focus position as well as the characteristics of the imaging system.

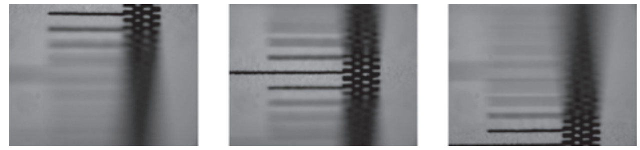


Figure 1: Sample images from a focal stack using imaging system with small depth of field.

Utilizing the relationship between blur and the distance to the focus, conventional shape from focus methods are composed of mainly three steps (Figure 1). Firstly, a stack of images are captured while the focus of the imaging system is shifted with respect to the object. This is typically implemented through either mechanical scanning of the camera/sample, or motorized focus shifting with the lens. Secondly, a focus measure value is calculated for each pixel of every image in the stack to form a 3D focus measure cube, where two dimensions represent the transverse spatial coordinates corresponding to the camera pixels and one dimension denotes the axial shift coordinate. The focus measure value can be calculated with various algorithms to evaluate how well the underlying pixel is in focus. Last but not least, depth information of each pixel is retrieved based on its focus measure curve (Figure 2). Regardless of the focus measure algorithms, the focus measure values for a specific pixel at any axial locations within the measurement range typically forms a spiky Gaussian-shaped signal. A naive approach is to take the axial position with maximum focus measure value as the position of the object at this position. More sophisticated approaches involve fitting of the focus measure curve as well as optimization techniques, such as total variation regularization.

According to estimation theory, the uncertainty when trying to retrieve the position of a Gaussian-like peak signal is largely dependent on the width of the peak. Since a narrower peak leads to more accurate measurement, a shape from focus system typically aims to achieve

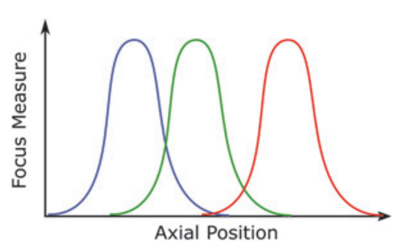


Figure 2: Sample focus measure curves for objects at different axial positions.

a depth of field as small as possible, by using optical systems with large aperture and longer focal length. Nevertheless, a narrower depth of field requires that the capturing of the focal stack has to be as dense as possible, which is very time consuming. This article aims to solve this contradiction.

2.2 Linear measurement model

Various real-world signals can be viewed as an n-dimensional vector $\mathbf{x} \in \mathbb{R}^n$, such as sound, image, etc. In a linear measurement model, each measurement of the target signal is a linear combination of all values in the vector $\mathbf{x}$. The complete measurements of the signal can be written as an m-dimensional vector $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ with an $m \times n$ measurement matrix A.

The ultimate goal of the linear measurement model, like any other measurement system, is to retrieve the signal $\mathbf{x}$ and the information it is carrying. Formulation of the linear measurement model as a linear system naturally leads to a classical problem of linear algebra: conditions for solving the equation $\mathbf{y} = A\mathbf{x}$. In this context, this question is equivalent to what kind of measurements are needed in order to recover the signal.

Although prevented by classical theory of linear algebra, recent developments in compressive sensing have shown that an underdetermined linear system can be uniquely solved provided sufficient prior knowledge [2]. In the case of compressive sensing, such prior knowledge refers to the assumption of sparsity. However, this is not the only possible prior knowledge. From a more general perspective, the underdetermined linear system with prior information represents a linear manifold learning problem where the prior information acts as the boundary of the manifold to be learned by its low-dimensional projection. The fundamental philosophy behind solutions of such problems is that the information embedded inside the high dimensional manifold is intrinsically of low dimension. In the case of compressive sensing, the unknown manifold is limited to hyperplanes spanned by a limited number of axes which corresponds to the sparsity assumption.

The significance of the linear measurement model to conventional SFF approaches is that the number of images required in the focal stack can be effectively compressed if each image can act as a linear combination of all originally required images in the focal stack. The focus measure stack can then be retrieved from the focus measure calculated from the compressed images. It should be noted that this is only possible when the focus measure operator is linear, which is seldom true for modern focus measure operators. Fortunately, most focus measure operators are composed of several sub-operators, and as long as there is at least one linear sub-operator before all nonlinear sub-operators, the reconstruction can be inserted (Figure 3). In other words, the first sub-operator applied on the compressed images must be linear.

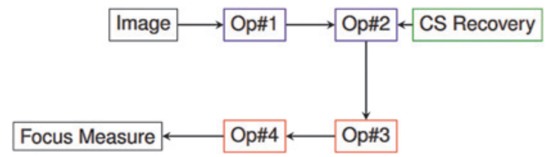


Figure 3: The compression and recovery steps must be added before all non-linear operators due to their linear nature. Blue: linear operators. Red: non-linear operators.

The algorithm for the reconstruction of the focus measure stack depends on the prior knowledge, i.e. the focus measure operator. On one hand, when the focus measure curve has a defined peak, recent compressive sensing algorithms can be incorporated for the recovery of the whole curve. On the other hand, if a training process is allowed or focus measure curves can be assumed, conventional methods like principle component analysis (PCA) can be applied in this scheme to yield the measurement/compressing matrix and the reconstruction matrix.

3 Proposed algorithm

To explain this idea in a concrete and clear manner, an exemplary algorithm is presented in this section. The schematic of the algorithm is illustrated in Figure 4.

The measurement matrix forming the compressed images and the reconstruction matrix for decompression are designed by a training process. In this process, conventional SFF procedures are implemented on a sample focal stack so that the focus measure curve for each pixel is calculated. All the focus measure curves are then assembled, with which PCA is conducted. The largest components are combined to construct the measurement matrix for the compressed images and the reconstruction matrix is simply the transpose of the measurement matrix. The focus measure curve can be reconstructed by multiplying the focus measure of the compressed images with the reconstruction matrix:

$$x_R = A^T y = A^T A x \tag{2}$$

where x is the original focus measure curve and A is the measurement matrix for compression.

The widely accepted modified Laplacian operator (LAPM) is selected for calculation of the focus measure [6].

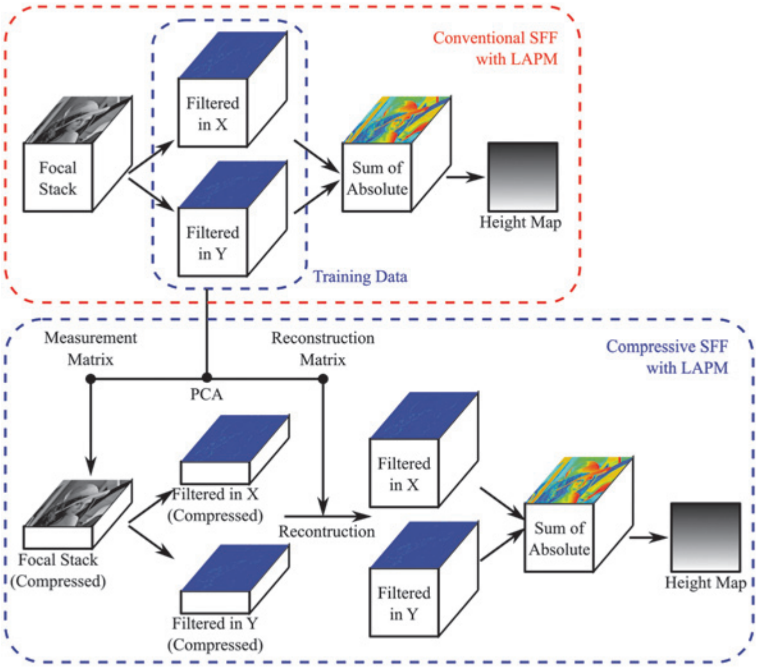


Figure 4: Schematic of compressive SFF with LAPM operator.

It consists of two sub-operators. Firstly, a one dimensional Laplacian filter is constructed as $f = (-1, 2, -1)$ and used to filter the image in both x and y directions respectively. Secondly, the absolute values of two filtered images are summed as the final focus measure value. Apparently the 1D filtering operation as a convolution is linear while taking the absolute value is non-linear. Therefore the training and reconstruction step must be inserted before taking the absolute value. From the recovered datacubes of filtering in X and Y directions, the final focus measure value can be computed through the sum of the two absolute values. Then for each pixel, the axial focus measure curve is smoothed before the maximum value is located to estimate the axial depth.

4 Simulation and discussion

4.1 Dataset construction

To demonstrate the applicability of the proposed algorithm, simulation is implemented in Matlab with a series of datasets synthetically generated through programs provided by Pertuz et al. in their survey study [7]. The generation of the focal stacks is based on a non-linear, shift variant model of defocus. All estimation results shown in this section are smoothed with a mean filter (window size = 5).

To investigate into the influence of training dataset on the compressive SFF (CSFF) result, two texture maps and two depthmaps are combined to form four different datasets (Figure 5). Texture #1 is a structured concentric pattern while texture #2 is a random pattern. Depthmap #1 is a linear ramp and depthmap #2 is part of a sphere. The four datasets #1–#4 are formed with combination of texture and depthmap in the following order: #1 and #1, #1 and #2, #2 and #1, #2 and #2.

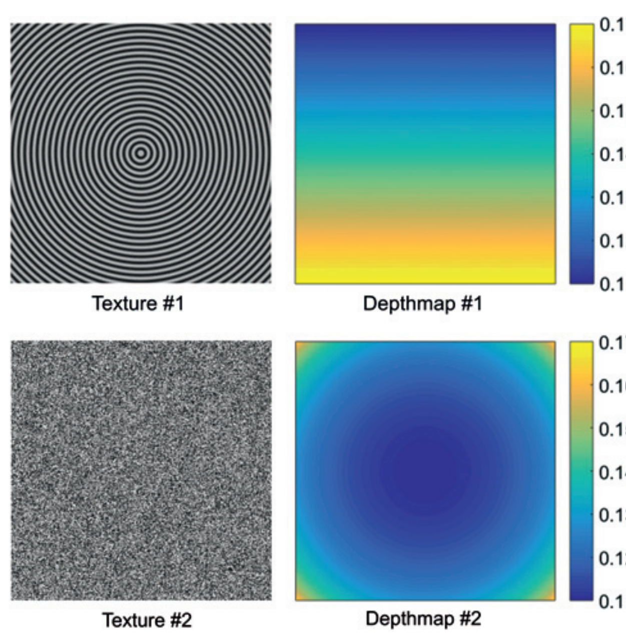


Figure 5: Textures and depthmaps used to sythesize focal stacks.

Table 1: RMS error showing influence of training set on testing result.

RMS Error ($\times 10^{-3}$)	Set #1	Set #2	Set #3	Set #4
No Training	1.69	1.74	4.92	3.25
Trained #1	NA	0.62	1.19	0.79
Trained #2	1.11	NA	2.98	0.72
Trained #3	4.23	2.40	NA	0.46
Trained #4	6.95	2.52	5.58	NA
Trained All Sets	3.65	1.93	1.32	0.54

4.2 Simulation result

Results of CSFF are compared with those of conventional SFF using the root-mean-square (RMS) error with respect to the ground-truth depthmaps, which are listed in Table 1. For the conventional method, a focal stack of 61 images is generated in each case and for the CSFF method the 61 images are compressed into 6 images. The row labeled as no training represents the conventional case whereas the other rows are labeled with their corresponding training set, which is used to generate the measurement matrix and the reconstruction matrix. It can be seen from Table 1 that the choice of training set has an influence on the testing result. In general, the compressive results are comparable to the conventional results but require much smaller numbers of compressively captured images. The test result of set #1 with training set #2 and the test result of set #4 with training set #1 are illustrated in Figure 6.

To investigate into the number of images needed for SFF, a series of focal stacks with different numbers of images is synthesized based on texture #1 and depthmap #1 (same combination as dataset #1 used in previous simulations). As expected for the conventional CSS scheme, when the number of images increases, the RMS error decreases, indicating better estimation result. This is due to the fact that the simulated imaging system for image synthesizing has a limited depth of field defined by the blurring kernel. When the step between two adjacent images is too large, the areas with depth in the interval between two focal planes will never get the chance to be imaged sharply and thus cannot be estimated robustly. With the conventional SFF scheme, the minimum number of images needed for robust estimation depends largely on the depth of field of the imaging system, which determines the width of the peak in the focus measure curve when using an operator like LAPM. Generally speaking, when the step size is larger than the width of the focus measure curve, artifacts will start to appear in the estimation result. The dependency of estimation accuracy on the number of images in the focal stack is illustrated by the blue solid curve in Figure 7.

On the contrary, in CSFF, regardless of the number of compressive images to be captured, each image acts as a linear combination of all focal planes within the measurement range, and thus contains information from all focal positions in an encoded manner. A training dataset based on texture #1 and depthmap #2 is synthesized with 100 images. The number of compressive images is solely determined by the number of largest principal components to be selected for the construction of the measurement matrix. As shown in Figure 7, CSFF allows much less

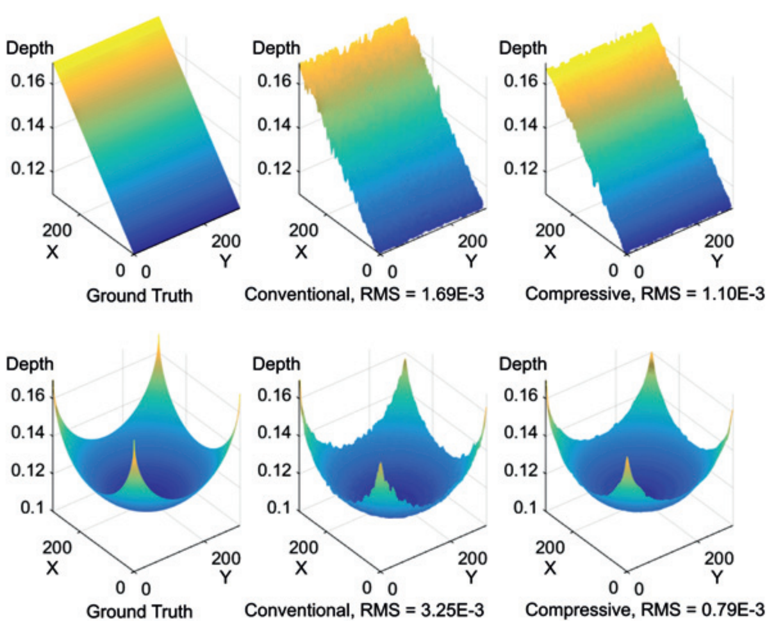


Figure 6: Depth estimation results with conventional SFF and compressive SFF. The upper row illustrates result of set #1 with training set #2 and the lower row illustrates result of set #4 with training set #1.

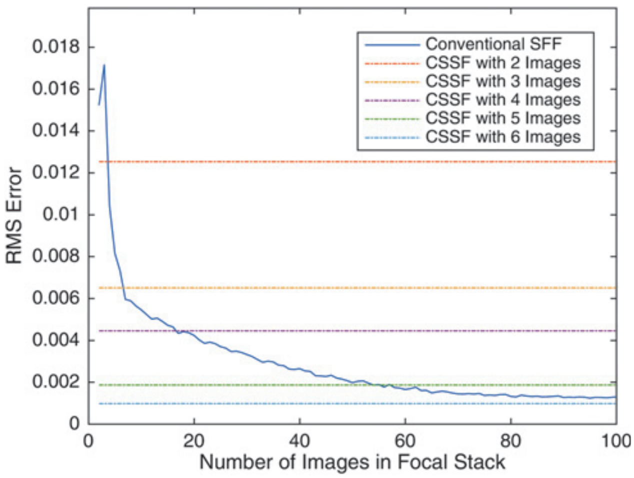


Figure 7: Dependency of estimation accuracy on the number of input images.

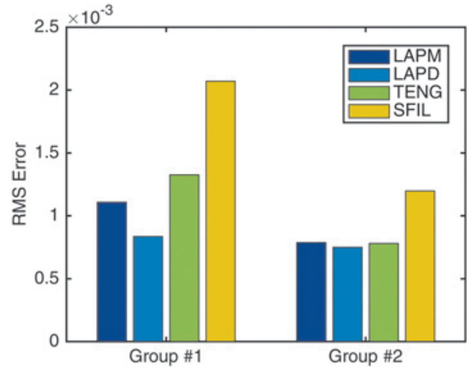


Figure 8: Comparison of CSFF using different focus measure algorithms. Group #1: dataset #1 with training set #2. Group #2: dataset #4 with training set #1.

images to be captured to achieve the same level of estimation accuracy as the conventional method. It should be noted that results presented above demonstrate the feasibility of the method only on the theoretical level with synthetic datasets. Further investigations with real datasets shall be made in future research. As the information contained in the largest principal components is related with the rank of the matrix, it is preferred to have a matrix with low rank. This means that the width of the focus measure curve should be larger, i.e. the imaging system should have a larger depth of field. However, as the width gets larger, the relative magnitude of the focus measure peak gets smaller, effectively reducing the SNR of the measurement. Therefore, a balance must be made between these two factors to generate the best estimation performance.

4.3 Applicability

To investigate into the applicability of the compressive approach, several other focus measure algorithms are implemented and simulated with synthetic datasets, including diagonal Laplacian algorithm (LAPD) [11], Tenegrad algorithm (TENG) [8] and steerable filters algorithm (SFIL) [5]. Similar to the modified Laplacian algorithm, the diagonal Laplacian algorithm also applies Laplacian operators to the captured images, but in two additional diagonal directions. Based on gradients of the image, the Tenegrad method applies Sobel filters to the image in both directions and then makes a squared sum. The steerable filters algorithm is a sophisticated modern algorithm that has attracted quite a lot of attention. The focus measure value is calculated using steerable filters in several directions, which are designed in quadrature pairs for better control over phase and orientation. The maximum of the filtered results is taken as the focus measure. Mathematical details regarding these algorithms can be found in the respective literature.

Two groups of tests are conducted using the compressive approach proposed previously with all four focus measure algorithms. For Group #1, dataset #1 is tested using the training result from dataset #2. For Group #2, dataset #4 is tested using the training result from dataset #1. All algorithms are modified so that the compression and reconstruction processes are inserted before any nonlinear operations. As shown in Figure 8, all four focus measure algorithms have provided similar results under the compressive scheme. Therefore, the compressive scheme is in general not very sensitive to the selection of focus measure algorithms as long as the linearity condition mentioned in Section 2.2 is satisfied. Nevertheless, for both groups of tests, the steerable filters algorithm has demonstrated noticeably worse results than the other three algorithms. As a more sophisticated algorithm, SFIL should in principle generate better results than the other three algorithms when applied in conventional SFF. However, due to its higher complexity, the added noise introduced by the compression and reconstruction might have more severe influence over the final focus measure calculation, leading to a worse overall result. This implies that the degeneration caused by the compression is possibly more severe for more complicated focus measure algorithms, which has to be taken into consideration when applying compressive shape from focus in practice.

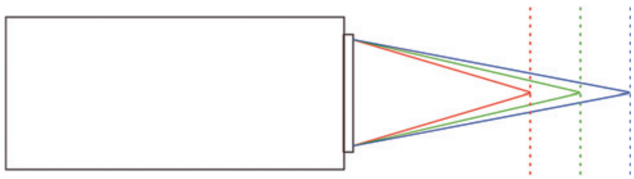


Figure 9: Possible hardware implementation of CSFF using hyperchromatic objective coupled with designed illumination spectra.

5 Conclusion

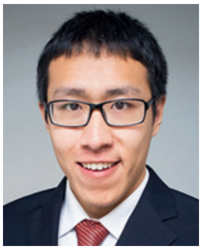
In this article, a novel scheme of compressive shape from focus is presented and simulated. Based on the linear measurement model, CSFF compressively captures several images, each as a linear combination of all possible focal planes within the measurement range. It has been shown in the simulation that the estimation error of CSFF is comparable to the conventional method using the same number of images as the number of images in the training set for CSFF. With datasets synthesized in this article, CSFF with 6 compressive images yields better performance than the conventional method with a focal stack of 100 images. Apart from LAPM, several other focus measure algorithms are also tested under the compressive approach, indicating wide applicability of the method. As research presented in this article represents an early stage of investigation into the compressive method, it should be noted that the feasibility of the method is only proven with synthetic datasets. To make a full investigation into this method with comparison to conventional methods, further research must be conducted with real data in the future.

Hardware implementations of this scheme could take many different forms. To begin with, auto-focus mechanisms in cameras can be modified so that the focal plane is shifted with a varying speed within one exposure according to the measurement matrix. Customized objectives based on liquid lenses can be developed to increase the speed of focus shifting. On the other hand, if the texture is colorless, hyperchromatic objectives can be designed and coupled with tunable illumination to achieve the same effect (Figure 9).

References

1. M. Asif and Tae-Sun Choi. Shape from focus using multilayer feedforward neural networks. *Image Processing, IEEE Transactions on*, 10(11):1670–1675, Nov 2001. ISSN 1057-7149. 10.1109/83.967395.
2. Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006. ISSN 1097-0312. 10.1002/cpa.20124.
3. M. T. Mahmood. Shape from focus by total variation. In *IVMSP Workshop, 2013 IEEE 11th*, pages 1–4, June 2013. 10.1109/IVMSPW.2013.6611940.
4. D. Marr and T. Poggio. Cooperative computation of stereo disparity. Technical report, Cambridge, MA, USA, 1976.
5. R. Minhas, A. A. Mohammed, Q. M. J. Wu, and M. A. Sid-Ahmed. *3D Shape from Focus and Depth Map Computation Using Steerable Filters*, pages 573–583. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-642-02611-9. 10.1007/978-3-642-02611-9_57.
6. S. K. Nayar and Y. Nakagawa. Shape from focus. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 16(8):824–831, Aug 1994. ISSN 0162-8828. 10.1109/34.308479.
7. S. Pertuz, D. Puig, and M. A. Garcia. Analysis of focus measure operators for shape-from-focus. *Pattern Recogn.*, 46(5):1415–1432, May 2013. ISSN 0031-3203. 10.1016/j.patcog.2012.11.011.
8. A. Santos, C. Ortiz de Solorzano, J. J. Vaquero, J. M. Peña, N. Malpica, and F. Del Pozo. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188(3):264–272, December 1997. 10.1046/j.1365-2818.1997.2630819.x.
9. Martin Schaffer, Marcus Grosse, Bastian Harendt, and Richard Kowarschik. High-speed three-dimensional shape measurements of objects with laser speckles and acousto-optical deflection. *Opt. Lett.*, 36(16):3097–3099, Aug 2011. 10.1364/OL.36.003097.
10. Murali Subbarao and Tae Choi. Accurate recovery of three-dimensional shape from image focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(3):266–274, March 1995. ISSN 0162-8828. 10.1109/34.368191.
11. A. Thelen, S. Frey, S. Hirsch, and P. Hering. Improvements in shape-from-focus for holographic reconstructions with regard to focus operators, neighborhood-size, and height value interpolation. *IEEE Transactions on Image Processing*, 18(1):151–157, Jan 2009. ISSN 1057-7149. 10.1109/TIP.2008.2007049.
12. C. T. Vu, T. D. Phan, and D. M. Chandler. S_3 : A spectral and spatial measure of local perceived sharpness in natural images. *Image Processing, IEEE Transactions on*, 21(3):934–945, March 2012. ISSN 1057-7149. 10.1109/TIP.2011.2169974.
13. R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8):690–706, Aug 1999. ISSN 0162-8828. 10.1109/34.784284.

Bionotes



Ding Luo
Karlsruher Institut für Technologie (KIT), Lehrstuhl für Interaktive Echtzeitsysteme, Karlsruhe, Germany
ding.luo@iosb.fraunhofer.de

Ding Luo received his master degree in Optics and Photonics from Karlsruhe Institute of Technology in 2014 and his bachelor degree in Information Engineering from Department of Optical Engineering, Zhejiang University in 2012. Since January 2015, he has joined Lehrstuhl IES, KIT in Germany as a scientific staff and has been working towards a PhD degree. He is currently working on adaptive chromatic measurement system. Ding Luo works in close cooperation with the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) in Karlsruhe.

Thomas Längle
Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB), Karlsruhe, Germany

Thomas Längle is associate professor at Karlsruhe Institute of Technology (KIT), Karlsruhe and the head of the business unit Visual Inspection Systems (SPR) at Fraunhofer IOSB in Karlsruhe, Germany. His research interests included different aspects of image processing and real-time algorithms for inspection systems. He also offers lectures in computer science at Karlsruhe Institute of Technology and initiates many possibilities for students to work on applied research.

Jürgen Beyerer
Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (IOSB), Karlsruhe, Germany

Jürgen Beyerer is the director of the Fraunhofer Institute of Optronics, System Technologies and Image Exploitation (IOSB) and the head of the Vision and Fusion Laboratory (IES) at the Faculty of Informatics, Karlsruhe Institute of Technology (KIT). His main fields of research are: Automated visual inspection and image processing, fusion of heterogeneous information sources, information theory, system theory, statistical methods and metrology.