MULTI-FOCUS IMAGING USING LOCAL FOCUS ESTIMATION AND MOSAICKING

Dmitry Fedorov, Baris Sumengen, B. S. Manjunath

Center for Bio-Image Informatics
Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106.
email: {fedorov,sumengen,manj}@ece.ucsb.edu
web: http://vision.ece.ucsb.edu, http://www.bioimage.ucsb.edu

ABSTRACT

We propose an algorithm to generate one multi-focus image from a set of images acquired at different focus settings. First images are registered to avoid large misalignments. Each image is tiled with overlapping neighborhoods. Then, for each region the tile that corresponds to the best focus is chosen to construct the multi-focus image. The overlapping tiles are then seamlessly mosaicked. Our approach is presented for images from optical microscopes and hand held consumer cameras, and demonstrates robustness to temporal changes and small misalignments. The implementation is computationally efficient and gives good results.

Index Terms— Focus, seamless mosaicking, microscopy, consumer cameras

1. INTRODUCTION

Modern optic systems carry several fundamental limitations. One of them is the low depth-of-field. Usually, certain objects at particular distance are focused while other objects are blurred to a degree depending on their distances from the camera (see Fig. 1). This problem is encountered in photography and microscopy. In microscopy, scientists use several images of the specimen with different focus parameters in order to see in detail the whole 3D structure. Meanwhile in digital microscopy there are solutions provided such as confocal microscope that collects light from only one focal plane and therefore discards all the scattered light from out-of-focus planes. On the other hand, confocal imaging is slower than the optical counterparts, and not suitable in many cases.

The imaging properties of an optical system depend on the acquisition parameters (focal length, focus, etc.) and on the distances to the objects imaged. Most of the previous work considered *depth from focus* (depth from defocus) problem, and the wavelet transform [1] and gradient domain [2, 3, 4, 5] solutions were proposed to estimate focus at each pixel of the

(a) Focus setting at 20cm   (b) Focus setting at 2m

Fig. 1. Images taken from the same location with different focus settings. Objects at particular distance are focused while other are blurred to a degree depending on their distances.

image. Many of these methods are either computationally expensive or susceptible to noise, acquisition quality, and temporal changes. We simplify this problem in order to estimate an image as a combination of maximum focus regions using a computationally efficient scheme. This offers a fast and robust solution to represent a 3D structure acquired by multiple images with different focus setting in 2D preserving its fine details. Our approach is robust to acquisition parameters and temporal changes. Since we blend the images using pixel data from the spatial domain (as opposed to fusing the information in a transform domain), the resulting images have fewer artifacts. The results presented for microscopy and hand held consumer cameras demonstrate good quality.

This paper is structured as follows. We describe proposed algorithm in section 2. First pre-registration algorithm is explained in section 2.1. Then the focus measure is presented in Section 2.2 and in section 2.3 construction of multi-focus image is described. Section 3 shows experimental results. We conclude in section 4.

2. MULTI-FOCUS IMAGING ALGORITHM

We use a series of images with different focus setting in order to generate one multi-focus image. Our algorithm accepts slightly miss-aligned image sets acquired with hand-held cameras. This would require a pre-registration step. During pre-registration input images are mapped into one common plane. Once aligned, input images are divided into small overlapping regions and for each position one region with the

best focus is estimated. Estimated regions are then seamlessly mosaicked back together and the result is rendered.

2.1. Pre registration

In order to use our method with hand-held consumer cameras, the input images obtained by manually varying the focus parameter must be pre-registered. The transformation model $x'_i = Hx_i$ that we have chosen is a planar homography, where $H \in \mathbb{R}^{3 \times 3}$ and has 8 degrees of freedom. The registration problem in presence of parallax is very hard by itself and therefore we allow only minor perturbations between images. Although, an homography can only model view changes of planar scenes the focus method is robust enough to handle small deviations from the purely planar scenario. Matched tie points are used to estimate the parameters of the model by the Normalized Direct Linear Transformation (DLT) algorithm [8]. First candidate locations are extracted as uniformly distributed local maximas over condition surface [6]. Then for each of these locations the point descriptors are extracted. Considering small perturbations among input images, our choice privileges the descriptors that can be computed efficiently. We use small circular windows whose intensity content is normalized and orientation is aligned with the average gradient orientation [7]. Preliminary matches of the tie points are established identifying the pairs with minimum distance in the descriptor space. Afterwards, the inevitable outliers are pruned off using RANSAC-like [8] algorithm. During the refinement procedure the algorithm selects the dominant plane which is then used for mapping. Finally, the mapping is done using bi-linear interpolation.

2.2. Focus Estimation

Let image I be an ideal sharp image and multiple images I_i are captured by changing the optics settings, where $i = 1, ..., K$ and K is the number of changes in the optics. The blurring effect in images I_i can be modeled as a convolution of the image I with an associated depth related Point Spread Function f_i of the optics: $I_i = f_i(x, y) * I(x, y)$, where $*$ denotes the two-dimensional convolution. It is demonstrated [2] that the estimation of the pixels's focus is related to studying a small neighborhood around that pixel. We relate this neighborhood to a small tile T used to divide input images and that will constitute a new resultant image. We then rephrase the image acquisition model as: $T_i = f_i(x, y) * T(x, y)$, where T denotes the tile from the ideally sharp image. Therefore we could state the problem as finding the tile that minimizes the effect of the PSF: $\hat{i} = \arg\min_i(T(x, y) - T_i(x, y))$. Thus, the result consists only of patches from existing images.

At this point the problem is narrowed down to a selection of a tile with the best focus. It was demonstrated that the value of the focus measure increases as the image sharpness or contrast increases [4], therefore the region with the maximum focus measure can be detected. Different methods were proposed in literature in order to analyze focus in a region, Fourier and wavelet [1] transforms were used to analyze fre-
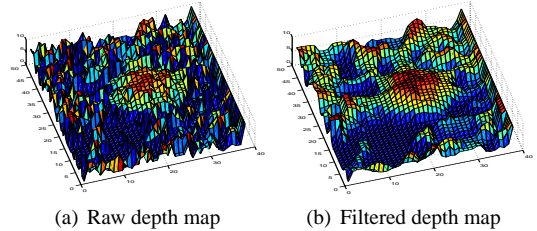

(a) Raw depth map   (b) Filtered depth map

Fig. 2. Raw (a) and filtered (b) depth maps for radioleria example, the z axis shows index of maximum focus tile T_i.

quency spectrum, several authors used gradient magnitude to obtain maximum focus measure [3, 5]. It was demonstrated that the focus measure operator should respond to high frequencies of image intensities and be maximal with maximum focus [3], furthermore it is possible to recover the PSF from the edges in the image [9]. Nayar also proposed to use second derivative to measure the focus [3] and used Modified Laplacian: $\nabla_M^2 I = \left| \frac{\partial^2 I}{\partial x^2} \right| + \left| \frac{\partial^2 I}{\partial y^2} \right|$. The proposed measure of focus for each pixel was a sum of Modified Laplacian values in a small window. We are interested in obtaining the focus measure for the whole tile and use measure F_i such as: $F_i = mean(\nabla_M^2 T_i) \cdot std(\nabla_M^2 T_i)$, where $mean$ and std denote average and standard deviation, and consequently the *maximum focus tile* $T_{\hat{i}}$ is identified by: $\hat{i} = \arg\max_i(F_i)$.

2.3. Constructing Multi-Focus Image

In this section the framework for image processing using local information is presented. The input images are divided into small overlapping tiles and then among the tiles that cover the same physical area we choose the one with best estimated focus. Selected tiles are then seamlessly mosaicked back together using multi resolution spline (MRS) technique [10]. The tile size is an important issue and should be comparable in size with the smallest object to be preserved locally. The minimum size for the tile is constrained by the use of MRS so that the image pyramid would still make sense.

The algorithm is divided into two steps. In the first step, the map of adjustment parameters for each tile is acquired. The adjustment parameter is the index of the tile with maximum focus. The map is generated sliding the tile-size window over the stack of input images by a certain "step". This "step" parameter is defined a priori and controls the amount of overlap which is usually half or quarter of the tile size. The generated map shows for each tile the image that focuses best that area, and therefore can be seen as a rough depth map. To guard against possible noise and enforce smoothness the maps are refined using filters such as median and gaussian. The obtained depth map is shown in Fig. 2 where z axis is index i of the tile T_i with maximum focus. The raw and filtered depth maps were obtained from radioleria example (Fig. 6).

In the second step, we render the resultant image by mosaicking consecutive tiles together. This process is done by rows where each row is constructed by consecutively blending neighboring tiles. In order to blend tiles we opt for multi-
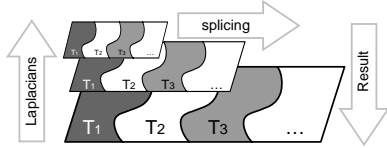
Fig. 3. Multi-resolution spline process for multiple tiles. Laplacians are spliced and recomposed into blended image.
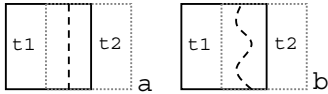

Fig. 4. Transition zones indicated by dashed lines (a) equidistant and (b) error minimization.

resolution spline technique [10] known to provide smooth blending yet preserving features located in the overlapping area. During this procedure, the images to be blended are first decomposed into a multi-resolution laplacian pyramid (Fig. 3). The pyramids are then spliced level by level, with each level being spliced using a weighted average over a transition zone. Then the blended image is obtained by reversely composing the spliced laplacian pyramid. Therefore the spline is matched to the scale of features and images are blended gradually without blurring finer image details. To improve speed using more memory the whole process can be done at once by assembling laplacians for the whole row or image and then reversing the whole structure.

The averaging transition zone can be easily defined as a line equidistant to borders of both tiles (Fig. 4(a)). The more elegant solution is to minimize the negative effect of the tile size with objects of curved structure by defining the transition zone between tiles as an error minimization problem (Fig. 4(b)). The error surface is defined as absolute value of the difference of gradients of overlapping tiles, T_a and T_b. $W = |\nabla_M^2 T_a - \nabla_M^2 T_b|$ This cost function provides splicing that avoids high difference areas and instead uses areas where gradients are of similar magnitudes. The approximate solution to this minimization problem is recently given by a computationally efficient graph-cut algorithm [11]. We define the graph where each node corresponds to a pixel in the overlapping area between the two tiles. The weight of the edge (p,q), where p and q are adjacent nodes, is defined by a cost function W. Source and sink links are also initialized for leftmost and right-most border pixels of overlapping area. Fig. 5 shows the comparison for MRS mosaicking with equidistant and error minimization transition zones. These results were generated using images presented in Fig. 1.

3. EXPERIMENTAL RESULTS

The performance of our algorithm is demonstrated on several examples. For natural scenes we use JPEG images acquired by Canon G5 camera with Bayer color sensor. For microscopy, we use 12 bit images acquired by a monochrome QImaging camera. The focus estimation is performed using

(a) Equidistant transition   (b) Error minimization transition

Fig. 5. Comparison of MRS mosaicking with equidistant (a) and error minimization (b) transitions. Notice wires in the top middle of the image and top-left extreme of the flower.

image intensities, and for the case of color images the YUV color space is used. The focus is estimated in the intensity (Y) channel and all three YUV channels are then mosaicked accordingly. Tile sizes in the range of 16 and up to 64 pixels were used. Fig. 6 shows a combination of nine images of radioleria digitally acquired by light microscope manually changing the focus setting. As seen, details of all nine images are captured in the generated multi-focus image (Fig. 6d). Fig. 7(c) shows a combination of four slightly misaligned focus images of integrated optical waveguide acquired on Scanning Electron Micrograph (SEM). The result of wavelet image fusion (Fig. 7(d)) from Matlab Wavelet Toolbox (MATI-FUS by Dr. Paul Zeeuw) demonstrates visible artifacts. The original images are focused on etched wafer surface behind waveguide (Fig. 7(a)) and on tilted waveguide sidewall (b). Fig. 8 shows two images of hydrant with focus at 20cm and 2m acquired by a consumer camera positioned on the tripod, there are small temporal changes due to wind. Automatic registration of these images returns 191 tie points estimating near unit transformation which is automatically ignored.

Additional data-sets and results are accessible from our web site: http://vision.ece.ucsb.edu/tileframework/.
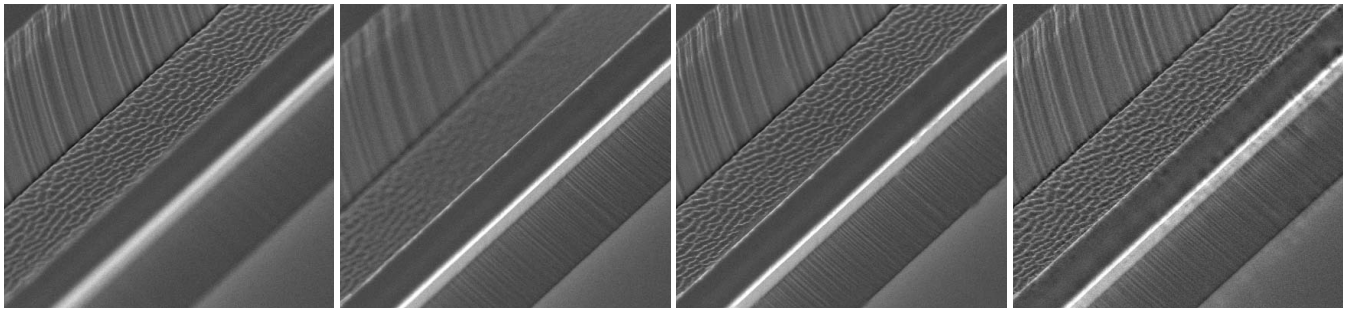
4. CONCLUSION AND FUTURE WORK

A tile-based multi-focus imaging technique is presented and its performance is demonstrated on a variety of images. Results show appealing visual quality for hand-held cameras and microscopy images. Our implementation is computationally efficient and can be parallelized since tile computations are independent. Our method is afflicted by temporal and geometric distortions between images, although, demonstrates better robustness than pixel-by-pixel methods. Currently we are investigating the application of this framework for local image enhancement and high dynamic range compression.

5. REFERENCES

[1] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor image fusion using the wavelet transform," in *first International Conference on Image Processing (ICIP'94)*, Nov 1994, vol. 1, pp. 51–55.
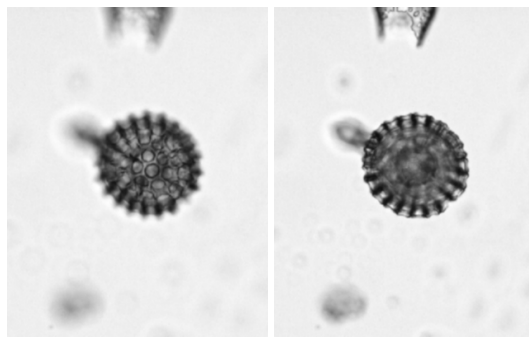
(a) Focus on etched wafer surface behind waveguide, left upper corner

(b) Focus on tilted waveguide sidewall, right lower corner

(c) Our multi-focus result

(d) Wavelet image fusion

Fig. 7. a-b) Two of the four slightly misaligned images of integrated optical waveguide acquired by Scanning Electron Micrograph (SEM). c) Our result. d) The result of wavelet image fusion.
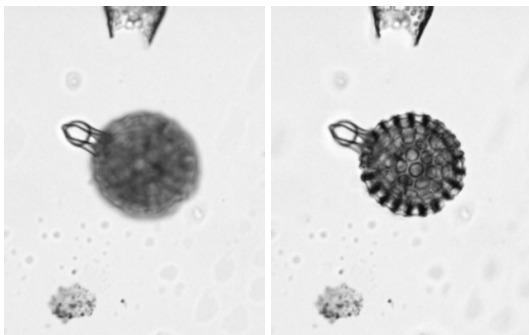

(a) Focus at 20cm, flowers are focused

(b) Focus at 2m, hydrant is focused

(c) Combined result

Fig. 8. Two images of hydrant with focus setting at 20cm (a) and 2m (b). Consumer camera positioned on the tripod, the images present small temporal changes due to wind. c) Combined resultant image.

(a) First image

(b) Fifth image

(c) Ninth image

(d) Generated multi-focus

Fig. 6. (a-c) three of 9 images of radioleria acquired by optical microscope and (d) combination.

[2] P. Favaro and S. Soatto, "A geometric approach to shape from defocus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 27, no. 3, pp. 406–417, 2005.

[3] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE T. Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[4] M. B. Ahmad and Tae-Sun Choi, "A heuristic approach for finding best focused shape.," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 15, no. 4, pp. 566–574, 2005.

[5] M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 8, pp. 864–870, 1998.

[6] C. S. Kenney, B. S. Manjunath, M. Zuliani, M. G. A. Hewer, and A. Van Nevel, "A condition number for point matching with application to registration and postregistration error estimation," *IEEE T. on Pattern Analysis and Mach. Intell.*, vol. 25, no. 11, pp. 1437–1454, Nov 2003.

[7] M. Zuliani, C. Kenney, and B. S. Manjunath, "A mathematical comparison of point detectors," in *Second IEEE Image and Video Registration Workshop (IVR)*, Jun 2004.

[8] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, second edition, 2004.

[9] M. Subbarao, T. Chio, and A. Nikzad, "Focusing techniques," *J. Optical Engineering*, vol. 32, no. 11, pp. 2824–2836, 1993.

[10] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics*, vol. 2, no. 4, pp. 217–236, 1983.

[11] Y. Boykov and V. Kolmogorov, "An experimental comparison of mincut/max-flow algorithms for energy minimization in vision," in *Energy Minimiz. Meth. in Comp. Vis. and Patt. Recog.*, 2001, pp. 359–374.