https://doi.org/10.2352/ISSN.2470-1173.2021.18.3DIA-063

An Automated Adaptive Focus Pipeline for Reflectance Transformation Imaging

David A. Lewis*, Marvin Nurit†, Hermine Chatoux‡, Fabrice Meriaudeau§, and Alamin Mansouri¶
Laboratoire Imagerie et Vision Artificielle, Université Bourgogne Franche-Comté,
Dijon, France
Email: *david_lewis@etu.u-bourgogne.fr, †marvin_nurit@etu.u-bourgogne.fr, ‡Hermine.Chatoux@u-bourgogne.fr,
§Fabrice.Meriaudeau@u-bourgogne.fr ¶alamin.mansouri@u-bourgogne.fr

***Abstract*—In this paper, we propose an automated adaptive focus pipeline for creating synthetic extended depth of field images using a reflectance transformation imaging (RTI) system. The pipeline proposed detects object regions at different depth levels relative to the camera's depth of field and collects a most focused image for each. These images are then run through a focus stacking algorithm to create an image where the focus of each pixel has been maximized for the given camera parameters, lighting conditions, and glare. As RTI is used for many cultural heritage imaging projects, automating this process provides high quality data by removing the need for many separate images focused on different regions of interest on the object. It also lowers the skill floor for this image collection process by reducing the amount of manual adjustments that need to be made for focus. Furthermore, this can help to minimize the amount of time that a sensitive cultural heritage object is outside of its ideal preservation environment.**

***Index Terms*—Adaptive focus, K-means clustering, Reflectance transformation imaging (RTI), Cultural heritage**

I. Introduction

Reflectance transformation imaging (RTI) is a method of imaging that is used to better understand the surface structure of an object through the analysis of computed surface normals. The system concept used for RTI incorporates a stationary birds-eye view camera and an illumination method that lights the subject from a variety of angles. A maximum resolution of each pixel is desired to better understand the reflective properties at each pixel for RTI. This also has the added benefit of providing a high quality image that can be used with the RTI output for a higher quality visualization of the object that could lead to greater understanding of the object itself.

The camera's depth of field can be decreased in order to increase the camera resolution, but this can introduce issues if the object being imaged is topographically complex. If this is the case, then not all points on the object will be in focus, undermining the attempt to increase resolution. If images are captured that are found to be focused at varying depth levels in the image, then they can be composited together to form a single image with an extended depth of field that possesses the high resolution benefits of a shallow depth of field acquisition.

Following this introduction, we present some of the related works that we are building off of. Then we will outline the algorithm that is used to drive to the image acquisition process as well as some results on both general focus variation and RTI imagery.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 813789.

II. Related Works

Estimating focus in an image as well as using multi-focus image fusion are well-defined and explored problems. However, these papers seem to have used curated sets of images, requiring the system operator to choose what input images are used. By combining these two processes, an automated image acquisition pipeline can be developed to remove this requirement.

The task of estimating focus can be re-framed as estimating the "sharpness" of an image. It's generally accepted in the field of image processing that "the condition of defocus results in attenuation of high spatial frequencies" [1]. Following this, a strong presence of high spatial frequencies implies a high level of focus. Edge analysis provides information on the presence of spatial frequencies in an image and thus an estimation of the level of focus. In the cited paper by Aslantas and Kurban, it was found that a computation of the sum-modified Laplacian was most effective for determining focus when attempting to fuse noise multi-focus images using discrete wavelet transforms when the fusion output image was compared to a "ground-truth" image with a wider depth of field than the fusion input images.

Yeo et al. [8] proposed a similar, yet different, autofocusing methodology for tissue microscopy and color imagery. They found that certain focus measures are more sensitive than others and that this could be used to create a two-stage focusing system. The system first performs a coarse search to locate the likely region of the best focused image which was followed by a finer search with the more sensitive focus measure. Furthermore, they proposed processing either a single color band, the first principal component, or a linear combination of the color bands of the image. We process grayscale input imagery that has been modified only by convolving with it a Gaussian blurring kernel that is of similar size to the focus operator window.

Huang and Jing [3] performed a methods survey concerning focus measures and how they can be used to fuse a multi-focus image dataset. The focus measures they tested were the variance of image gray levels, the energy of the image gradient, the Tenengrad method, the energy of the Laplacian of the image, the sum-modified Laplacian (SML), and an analysis of spatial frequency. When comparing the performance of these focus measures on input imagery, the images were split into smaller image blocks that were compared against one another before being used for the single output image.

Throughout the field, many variations and studies can be found on focus operators [7] and autofocusing methods, with many still concerned primarily with imagery from microscopes [2], [5]. One could surmise that a reason for this is due to the fact that more natural scenes are more complex than when images are captured in a controlled environment. Therefore, these approaches can still function as building blocks for our proposed pipeline since RTI systems also rely on a controlled environment. Factors that would normally need to be taken into account, such as lighting and backgrounds, are already parameters that need to be carefully considered for RTI.

III. Methods

The proposed pipeline consists of four main processes:

- a maximally focused image search,
- depth level segmentation,
- depth level focus optimization,
- and multi-focus image fusion.

The point of the first two processes is to investigate and enhance understanding of the “focus space” available for imaging. The term “focus space” is used to describe the range that the focus can be varied through the system, accomplished by moving the camera along the optical axis in this case. The latter two processes are run for each depth level detected by the segmentation process.

The first process takes the entire focus space and performs a search similar to a binary search, but splits the space into quadrants instead of halves. The reasoning behind this modification is not to increase precision, but to gather more imagery to better describe the space in the segmentation process. This search process takes in the entire focus space with the goal of finding a single point that results in a maximally focused image. The process of focus estimation consists of finding the mean of an image that has been processed with a focus measure operator, for which we chose to use the Laplacian operator.

While the system has captured all the input images while focusing on different planes of the object, it has perhaps not yet found images that can be used to best focus the various depth levels on the object. To get a better understanding of the object’s structure, each pixel in the scene can be treated as a feature in K-means clustering, with the same pixel in each image being a new sample. K-means clustering is used to cluster similar features together and therefore can be used to group together pixels that behave similarly across the sampled focus space.

K-means clustering typically initializes the center points for the clusters randomly, so K-means++ is used to select center points that are more informed by the present data and that are also reproducible. To choose these centers, K-means clustering still requires the desired number of clusters as an input. Since this is unknown, the elbow method can be used to analyze the within-cluster sum of squares, a measure of variability for each resulting cluster. This technique is used in data science to manually identify when an increase in clusters is no longer going to provide a significant increase in information. This method can be further augmented with the Kneedle method [6]. This method is capable of automatically traversing a set of data to determine the location of any elbows. This allows for the selection of an optimal number of clusters to use for K-means after a certain number of iterations has been reached.

The output regions from the K-means algorithm are representative of the depth levels on the object that hold important information. These regions are passed into a process similar to the original maximally focused image search, except that now we are optimizing just for the region of interest. This allows us to obtain images that are maximally focused for each depth level. These images are then passed into the multi-focus image fusion process, where all the images are combined into one with an artificially extended depth of field. This is achieved by comparing all the processed images with each other and selecting at each pixel the image that has the highest estimated focus value. The relevant image’s pixel value is placed at the appropriate coordinates in the output image.

After going through this process and collecting a number of images, it might seem reasonable to use all the acquired images as an input into the compositing process. Since this pipeline is developed with an RTI system in mind, however, this would then amount to a large number of images collected at what are typically numerous light positions. Since one can work under the assumption that the surface geometry does not change relative to the camera in a photometric stereo system, the list of determined best focus positions can be used to task the camera at all subsequent light positions. Therefore while the proposed pipeline may introduce a greater computational cost upfront, it can help save time and increase efficiency throughout the rest of the RTI acquisition.

IV. Proof of Concept with Pre-captured Images

A. Experiments

Before implementation into the system, the proposed pipeline was tested on a pre-captured set of images at various camera positions. The primary differences between this proof of concept and the implementation with the imaging system is that the former uses a large discrete image space while the latter is tasked with traversing a continuous image space until a threshold of precision is passed. The image sets used with this proof of concept were captured using our RTI system.

The image sets used varied between smaller sets ($\approx$ 25 images) or larger sets ($\approx$ 100 images) to test how the search would behave with different levels of simulated movement precision.

Despite this discrete space, the focus optimization search acts in much the same way as proposed for the continuous space, just with a preset level of precision.

(a) Image index 3/10

(b) Image index 6/10

(c) Image index 8/10

(d) Image index 10/10

(e) Image index 10/10

Fig. 1: Chosen images from focus stack and resulting composited image with extended depth of field

B. Results

As part of the validation of the proposed algorithm, a large number of sample datasets was sought. A dataset generated by Li et al. [4] was used for the wide variety of subjects and scenes collected with a varying focus. A single chosen example is presented here.

The images that were input into the focus stack can be seen in Figures 1 (a-d) as an example of the depth of field. Figure 1e is the output composite image created after traversing the input image set and choosing a set of images to composite. Visually, the image is focused at the different depth levels compared to any single input image. There are halo artifacts present around some edges, but the presence of this could be contributed to some combination of the relatively low image resolution and the kernel size used when evaluating the focus measure.

V. Real-Time Acquisition with RTI System

A. Experiments

The RTI system used for testing and applying this pipeline was developed in-house. It consists of a monochromatic Allied Vision Manta camera (model G1236) and a Qioptic Optem FUSION micro-imaging system. The latter item is a modular system that allows the user to adjust the focus using a motor or by interfacing with it programmatically. This allows one to quickly swap out various optical components while also providing a method of easily controlling the imaging system during a series of acquisitions.

B. Results

Possibly the most unique characteristic of an RTI acquisition is the variation of the angle of illumination while being able to leverage the consistency of a single viewpoint. With this in mind, it is important to look at the behavior of the proposed adaptive focus pipeline at a variety of illumination angles.

One of the subjects used for testing was a 1000 Colombian peso coin. While it does not have significant differences in surface height, it possesses both surface textures that can be detected by the proposed algorithm and reflective properties that are of interest in RTI applications. These adaptive focus acquisitions were compared against an alternative single image acquisition with a wider depth of field.

When utilized in the context of RTI, the shallow depth of field can provide a significant increase in image detail even on seemingly less complex objects such as this coin. The grazing angle illumination seen in Figure 4b provides no visual clues about the object when using a wide depth of field. When performing the same acquisition with a shallow depth of field, as seen in Figure 4c, and using the proposed pipeline to generate a composite image, some structure of the coin is revealed. This might not be a visually interesting image on its own, but taken in the context of RTI this can increase the range of possible information that can be used to augment the resulting reflectance understanding.

It can be seen in Figures 2 and 3 that the algorithm behaves as expected. As the elevation angle of the light changes from a minimum of 0° to a maximum of 90°, we would expect the increase in focus to reach a maximum at the halfway mark because most of the surface geometry is more perpendicular to the optical axis. This means that a point near the halfway point of the available range, there should be a minimal amount of glare while the illumination of the image features is maximized as the light is not at a grazing angle.

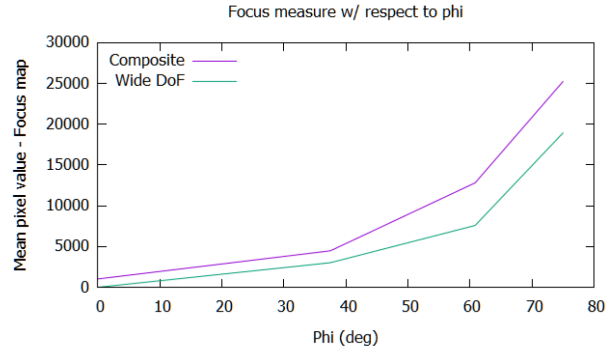


Fig. 2: Focus measure with respect to elevation angle

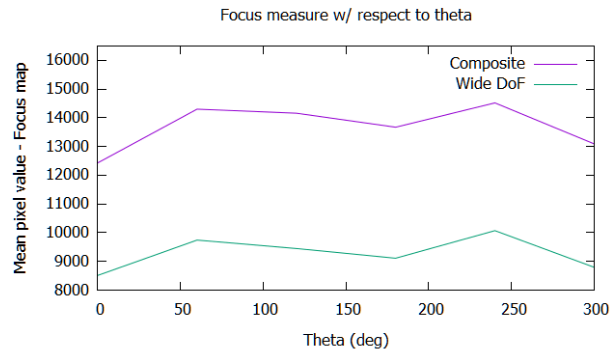
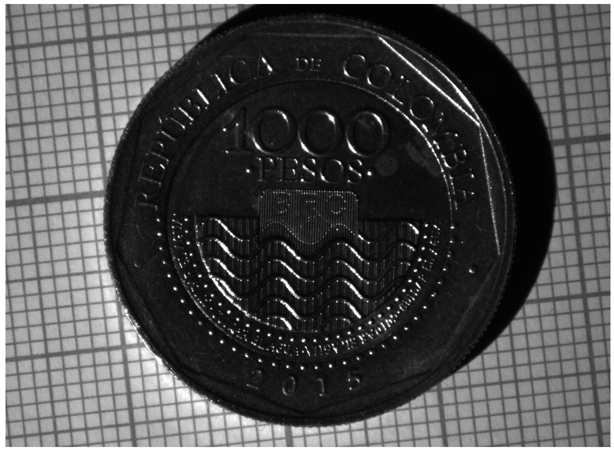


Fig. 3: Focus measure with respect to azimuth angle

VI. Limitations and Sources of Error

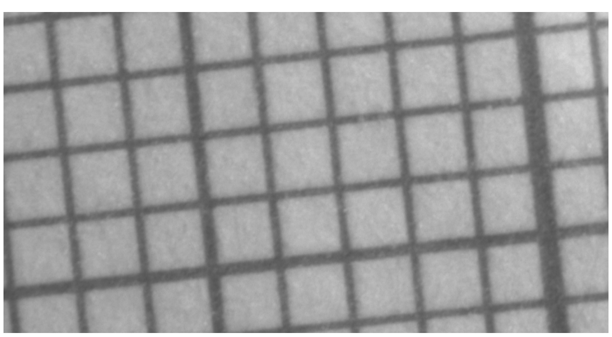
Spatial correlation is an important consideration for this pipeline. When making adjustments to the imaging system, it is possible to alter the object distance, which can cause the alignment of two images to be offset. This algorithm relies on the assumption that pixel neighborhoods are consistent across the image stack, and this misalignment can break this assumption. To this end, an attempt to align the imagery is made. For image stacks with significant blurring however, it can be difficult to accurately compute the homography between the two views. This means that to overcome this issue, either the difference between the two views must be already known and used to help compute the homography or that there has to be some ground truth map that can be used to correlate the images that are focused differently, such as an image taken with a wide depth of field. The current implementation seems to experience some issues with alignment as can be seen in the comparison in Figures 4d and 4e. This could also be due to the current compositing process because a relatively simple image compositing method is applied. Further analysis needs



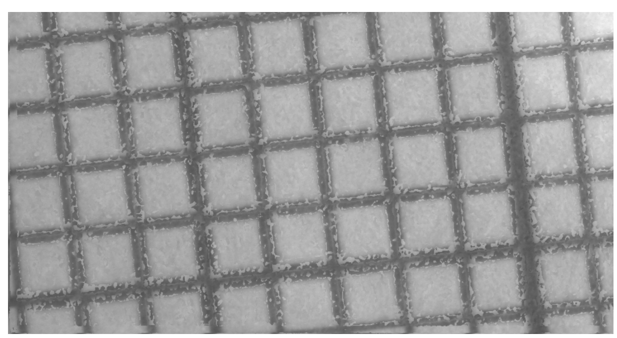
(a) Wide depth of field, at ($\theta = 120°$, $\phi = 37.5°$)

(b) Wide depth of field image with light position of ($\theta = 0°$, $\phi = 0°$)

(c) Composite image with light position of ($\theta = 0°$, $\phi = 0°$)



(d) Single focus stack input image with light position of ($\theta = 120°$, $\phi = 37.5°$)



(e) Composite image zoom with light position of ($\theta = 120°$, $\phi = 37.5°$)

Fig. 4: Chosen images from focus stack and resulting composited image with extended depth of field

to be conducted to determine the source(s) of these image artifacts. While this paper has not proposed an improvement on any methods of image compositing, this current limitation in the system must be discussed as there is the possibility that it introduces a not insignificant level of error. With the current results seen in Figures 2 and 3, it can be hypothesized that the error introduced isn't reliant on the amount of detail in an image but instead is a more consistent noise level found in the image compositing process. However, it is most likely that this noise can be contributed primarily to the captured dataset, as Figure 1 shows no evidence of this level of error. Nonetheless, the results are still promising due to the consistent behavior seen in Figures 2 and 3.

VII. Conclusion

In this paper, we have proposed an acquisition pipeline that seeks to adaptively focus on depth levels of significant information when imaging topographically complex objects. The main feature of the pipeline works as intended, determining different depth levels in a scene that contain high amounts of spatial frequency information. Given a space within which the focus of the scene changes, multiple points can be determined in this space that, when composited together, aim to have a maximum number of pixels in focus. It has also been shown that this method of image acquisition can augment RTI by increasing the range of light positions that generate usable imagery. Furthermore, the proposed pipeline generally behaves expectedly and stably as the light position changes relative to the subject being imaged, increasing the focus while retaining the relative image detail. Going forward, this work will seek to incorporate more robust methodology for multi-focus image compositing to improve output. For further incorporating this method with RTI, the variables associated with the light position will be taken into consideration in an attempt to further understanding of the subject being imaged.

References

[1] V Aslantas and Rifat Kurban. A comparison of criterion functions for fusion of multi-focus noisy images. *Optics Communications*, 282(16):3231–3242, 2009.

[2] Josue´ A´lvarez-Borrego. Fast autofocus algorithm for automated microscopes. *Optical Engineering*, 44(6):063601, June 2005.

[3] Wei Huang and Zhongliang Jing. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28(4):493–500, March 2007.

[4] Nianyi Li, Jinwei Ye, Yu Ji, Haibin Ling, and Jingyi Yu. Saliency detection on light field. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2806–2813, 2014.

[5] A. Santos, C. Ortiz De Solórzano, 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.

[6] Ville Satopaa, Jeannie Albrecht, David Irwin, and Barath Raghavan. Finding a" kneedle" in a haystack: Detecting knee points in system behavior. In *2011 31st international conference on distributed computing systems workshops*, pages 166–171. IEEE, 2011.

[7] Loren Shih. Autofocus survey: A comparison of algorithms. *Digital Photography III*, 6502:65020B, 2007.

[8] T T E Yeo, S H Ong, R Sinniah, and Jayasooriah. Autofocusing for tissue microscopy. page 11.