Machine Learn All Possible Photos

Akila de Silva
UCSC
audesilv@ucsc.edu

Fahim Hasan Khan
UCSC
fkhan4@ucsc.edu

Issei Mori
UCSC
imori@ucsc.edu

Abstract

Generating all possible combinations of images by synthetically applying the effects of various capture parameters of the camera can be a useful tool in the domain of image processing. The statistics of naturally occurring images lie in a low dimensional space. It is very unlikely to have a pure red pixel immediately adjacent to a pure blue pixel, and normally there should be a pixel of some color in between red and blue. The information can be used in a training set of images to make a prediction about future images and even generate them. Any methods which attempt to learn this subspace and project the desired result from the available data has a lot of application in real life and research fields. While the solution can be as simple as PCA, or vector quantization, or as complex as the state of the art recent machine learning (ML) technique. In this work, we propose a novel approach for predicting all possible variations of photos that can be captured using a professional DSLR camera by all the possible combinations of various capture parameters such as focus or aperture size. While our approach is based on machine learning techniques, we compared our results with the previous ML and non-ML based techniques as well as explored a simple approach based on curve fitting.

1. Introduction

(Motivation) - Although there are different types of digital and analog cameras for photography are available, all of these devices work based on the same principle of sensing lights passing through one or more lenses to reach the photosensitive material of the camera. So, natural images are captured based on the physical properties of light using a camera by adjusting a set of capture parameters. As the statistics of naturally occurring images lie in a low dimensional space, it is possible to generate all possible combinations of images by synthetically applying the effects of various capture parameters. It is very unlikely to have a pure red pixel immediately adjacent to a pure blue pixel, and normally, there should be a reddish blueish pixel in between. So there is a naturally occurring pattern of these pixels, and the information in a training set of images can be used to predict future images. Synthetic image generator such that can be a useful tool in the domain of image processing. The motivation of our work is to come up with a technique for predicting all possible variations of photos that can be captured using a professional DSLR camera by all the possible combinations of capture parameters such as focus or aperture size.

(a) Our prediction (b) Ground truth

Figure 1. (a): The predicted image of f/3.5 using our method with only 4 sample images taken with different apertures. (b): The ground truth image of f/3.5. Notice that our method produces an image that looks similar to the ground truth image without noticeable artifacts

(Challenge) - The primary challenge of the work is finding the best approach for generating all possible combinations of photos by mimicking the effects of the capture parameters in a physical DSLR camera. We need to ensure that the generated photos are realistic and natural and very close to the photos taken using the same capture parameters using a real camera. There are many potential solutions for doing this, which can be as simple as PCA, or vector quantization, or as complex as the latest state of the art machine learning solution. The first step towards tackling this challenge is to do some initial analysis of some image datasets and understand the statistical properties of them. The next step towards this challenge is to determine the most feasible solution and compare them with the previous and related work to select the optimal one.

(Obvious solutions) - There is no obvious solution to this problem, as an exactly similar problem was never addressed in any of the previous works by considering all of the capture parameters. However, from the initial assessment of the problem, the straight forward solution would be gener-

Figure 2. Our two datasets created with photos taken by changing the fouus and aperture size of two different DSLR cameras

ating images using curve fitting. Other obvious solutions are either an approach based on statistical analysis such as PCA, or vector quantization, or an ML-based solution, such as using deep neural networks. (Why obvious won't work) - Our experimentation using the linear and cubic curve-fitting approaches provided us some results, which has "halo" like artifacts. For the other statistical and ML-based approaches, more experimentations are necessary to determine if these are capable of using all the capture parameters to generate images.

(What we did) - As the first step, we started analyzing some datasets to discover the pattern and relationship among the changing pixels with the changing of different capture parameters. Then we generated some images by using linear and piece-wise cubic polynomial based curve-fitting methods to analyze the limitations of this approach. While we used regression analysis as a preliminary ML-based approach, we are planning to use a GAN based approach for our final results. (Results in brief) - The initial results presented here are some analyses of patterns of pixel intensity changes with changing focus and aperture size as capture parameters as well as some generated images using curve-fitting and regression analysis.

(Contribution Statements) - The main contribution of our work is a GAN based approach for generating all the possible images that can be taken with all possible setup of capture parameters (initially focus and aperture size) of a camera just by using a few photos. For that purpose, we created several datasets of images using a set of capture parameters by varying focus and aperture size. We compared the generated images from our approach with images generated using the curve fitting and regression analysis based approaches as well as compared our work with the previous works. Finally, we provided some guidelines about the extension of this work and potential future works.

2. Related Works

Even though we approached the problem of generating all possible images from a novel and different perspective, there are a few related works that utilized general and ML-based techniques to solve some similar problems.

General Techniques: A wide number of general techniques for image corrections were proposed in the last few decades, and we are only discussing some of the comparatively recent and closely related ones. A paper on synthetic Defocus and Look-Ahead Autofocus for Casual Videography introduce a technique to refocus on the main object in the scene [7]. A stack of images, captured sometimes using different aperture and focal length, are shown to be used in ordered to generate an desired image. Agarwala et al. presented Interactive digital photomontage and introduced a method of choosing a different image from a stack of images for individual faces to capture a group photo where everyone has desired face [1]. Hasinoff et al. presented Light-efficient photography and introduced a method of finding optimal camera settings for a series of shooting in order to capture a scene with a given aperture at a given exposure level in the shortest amount of time possible [12]. Another paper [4] presents a theory of focal stack compositing, and algorithms for computing images with extended depth of field, shallower depth of field than the lens aperture naturally provides, or even freeform or non-physical depth of field. This work shows that while these composites are subject to halo artifacts, there is a principled methodology for avoiding these artifacts—by feathering a slice selection map according to certain rules before computing the composite image. A followup to this work presents dynamic image stacks, an interactive and exploratory image viewer [5]. It explores what photography can become when the constraint by the implicit association between a display pixel and a static RGB value is relaxed. This proposed system first captures a burst of images with varying capture parameters. Then, in response to simple touch gestures on the image, the interactive viewer displays the best available image at the user's focus of attention. Exposure, focus, or white balance may be slightly compromised in the periphery, but the image parameters are optimal at the selected location.

ML based techniques: Several recent deep machine

Figure 3. Photos captured from the same scenario by a DLSR camera using different aperture sizes

learning approaches tried to solve similar problems, which include image generations. Many of these works are based on the concept of Generative adversarial nets (GAN) introduced by Goodfellow, Ian, et al. in [2], the first paper written on generative adversarial networks. A followup work [8] builds upon the [2] by introducing how to apply a condition on both the generator and the discriminator networks. Isola, Phillip, et al. presented a technique for Image-to-image translation with conditional adversarial networks. This paper investigates how conditional GAN [8] can be used in image-to-image translation problems [3]. A method called RefocusGAN was introduced based on a deblur-then-reblur approach to single image refocusing. Although, a number of special cameras, such as a Light Field camera, allows users to modify the depth of field at the post-processing stage [6] and [9], these cameras are not widely available. Their paper shows light field cameras can be learned using ML techniques to allow users to refocus images captured using a regular camera. They trained conditional adversarial networks for deblurring and refocusing using wide-aperture images created from light-fields. This approach achieved generic free-form refocusing over a single image by appropriately conditioning the networks with a focus measure, an in-focus image, and a refocus control parameter [11]. There is another paper on the Machine Learning approach for Non-blind Image Deconvolution, which introduces a technique to recover a sharp image from a blurry image, replacing a previous method using the Fourier domain [13]. A work by Wang, Ting-Chun, et al. presents a novel method for synthesizing high-resolution photo-realistic images from semantic label maps using conditional generative adversarial networks (conditional GANs). The paper demonstrated that an image-to-image synthesis pipeline could be extended to produce diverse outputs and enable interactive image manipulation given appropriate training input-output pairs [14]. Xian, Wenqi, et al. presented Texturegan for controlling deep image synthesis with texture patches. This paper investigates deep image synthesis and proposes an algorithm that can generate plausible images that are faithful to user controls. Although the paper is not exactly related to this project, its pipeline seems interesting to read more details [15]. GauGAN was presented in SIGGRAPH2019, which is a GAN-based image synthesis model that can generate photo-realistic images given an input semantic layout [10]. GauGAN is one of the recent research efforts in advancing GANs for real-time image rendering.

3. Results

Most of the related works addressed image generation just by using one capture parameter, such as only focus or aperture size. However, our work involves using all the available capture parameters to generate all possible variations of images by tuning them up.

4. Technology/Methods

While a few options can be used for generating all possible combinations of photos by mimicking the effects of changing the capture parameters focus and aperture size in a physical DSLR camera, determining the best approach is challenging. The generated photos need to be the same or at least very closely similar to the photos taken using the same capture parameters using a real camera. They should also look realistic and natural in human eyes. As the first step for addressing this challenge, we created some datasets by varying the focus and aperture size of DSLR cameras. We used two different models of DSLR cameras manufactured by Canon (EOS M50) and Nikon (??) two create our datasets and understanding the effect of using different devices with our approach (FIGURE). As any changes in lighting can affect the consistency of the images captured by varying focus and aperture size in the same dataset, the

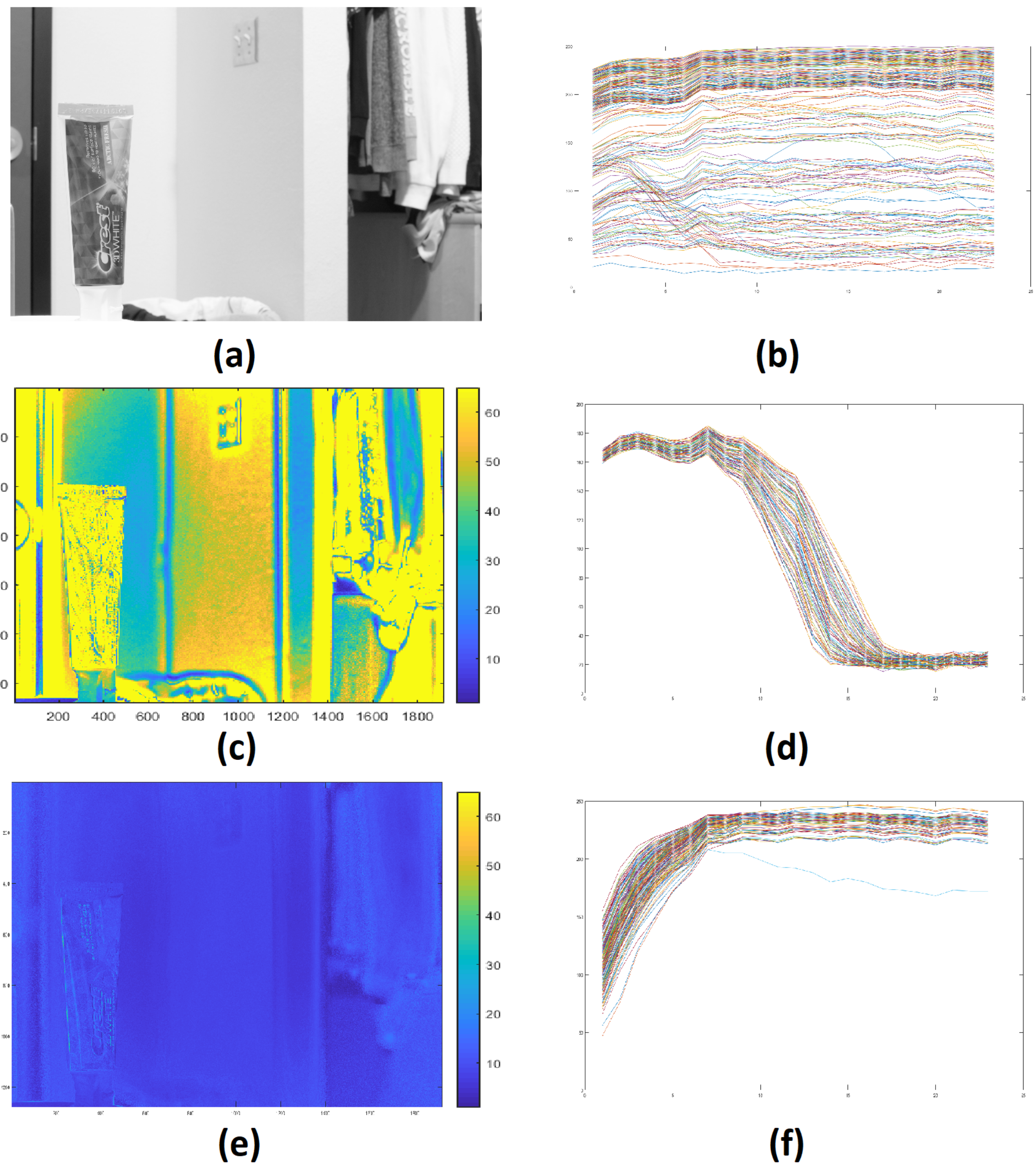


Figure 4. (a) A sample image from one of the datasets which contain twenty-three successive photos or frames taken by changing the aperture size of the DSLR camera. (b) The changes of the grayscale pixel intensity values in each successive frames. The horizontal axis represents the sequence number of the frames, while the vertical axis presents the sequential number of the pixels. Each line shows the gradual intensity change of a single-pixel. (c) The visual representation of the variance of each pixel with respect to the changes in the sequential frames. (d) The changes of the grayscale intensity of only pixels with higher variance than a predefined threshold value. (e) The visual representation of the maximum gradient values for each pixel with respect to the changes in the sequential frames. (f) The changes of the grayscale intensity of only pixels with higher maximum gradient than a predefined threshold value.

photos are taken for scenarios from indoor locations with controlled lighting and using a fully automated capturing software to maintain consistency. In order to minimize the noise, all the images are captured with a low ISO value, 100. We captured a total of 506 images per scene, F1.8 to F22, 22 distinct aperture values and 23 distinct focal length. Each image with a specific aperture value are captured with a corresponding shutter speed, in order to maintain the con-

(a) (b)

Figure 5. Some results of generating images by using linear curve fitting are presented here. Column (a) has four samples of original images, that can be considered as ground truth. Column (b) presents the four samples of generated images respective to the images on the left side using linear curve fitting approach.

sistency of the exposure. Then these captured images are re-scaled by a factor of 4, using a simple color averaging, in order to eliminate any noise produced by default. Then we did some initial analysis of our image datasets to understand the pattern and statistical properties of the changed pixels. The objective of this analysis is to discover the pattern and relationship among the changing pixels with varying focus and aperture size.

Figure 6. Some results of generating images by using a cubic polynomial curve fitting are presented here

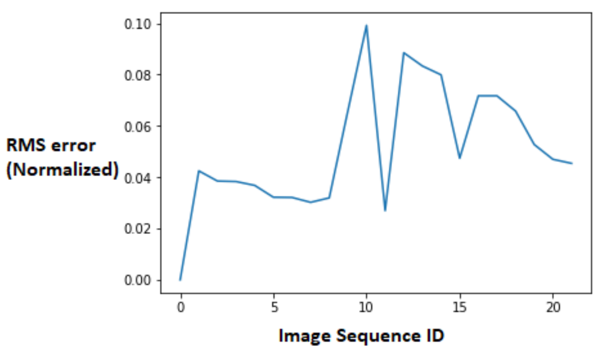


Figure 7. The Root Mean Square (RMS) Error for the generated images

From the initial assessment of the problem, the straight forward solution would be generating images using curve fitting. So, we generated some images by using curve-fitting by linear and piecewise cubic polynomials approaches and compared the generated images with the original images. While some of the generated images are very similar to the original image and look natural and realistic, there are some artifacts in some of the generated images. Examining the pixel-based analysis and the resulting photos generated, we can demonstrate the limitations of this approach. Some of the results generated using the curve fitting is presented in FIGURE 5 and explained in the results section.

The other approach we used in this paper is an ML-based solution, such as using deep neural networks. Initially, we used regression analysis as a preliminary ML-based approach for generating some images. However, our objective is to use a modified GAN [2] based approach for generating the images for all possible focus and aperture settings. The traditional GAN framework has one network called the "Generator" for generating image sequences, and another network called the "Discriminator" for evaluating the generated images. The generator learns to map from the training dataset to a particular data distribution of interest representing the intended movement, while the discriminator distinguishes between instances from the actual data distribution and results produced by the generator. The generator's training objective is to convince the discriminator to pass the generated images as real ones. Training the discriminator involves presenting it with image samples from the input dataset until it reaches some level of accuracy. The discriminator attempts to distinguish between fake samples produced by the generator and real ones sampled from the training data, which trains the generator to create more realistic images.GAN framework has one network called the "Generator" for generating image sequences, and another network called the "Discriminator" for evaluating the generated images [2]. The generator learns to map from the training dataset to a particular data distribution of interest representing the intended movement, while the discriminator distinguishes between instances from the actual data distribution and results produced by the generator. The generator's training objective is to convince the discriminator to pass the generated images as real ones. Training the dis-

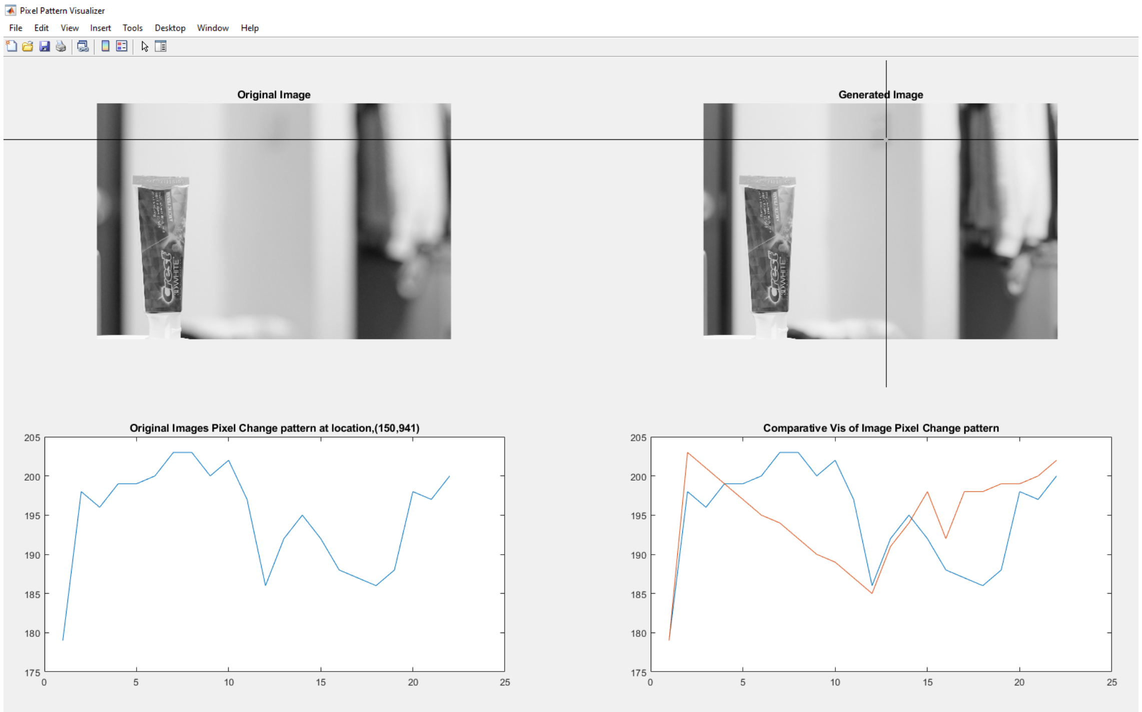
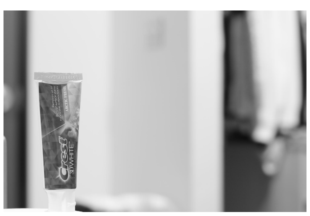
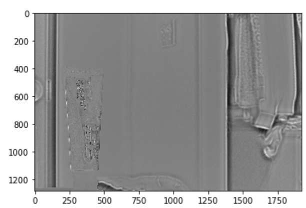


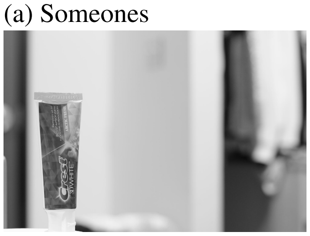
Figure 8. GUI of the visualization tool for comparing the image pixel change pattern in the original and generated images



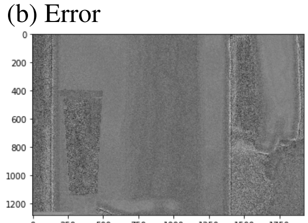
(a) Someones



(b) Error



(c) Ours



(d) Error

Figure 9. Left: The predicted image of f/3.5 using our method. Right: The ground truth image of f/3.5.

criminator involves presenting it with image samples from the input dataset until it reaches some level of accuracy. The discriminator attempts to distinguish between fake samples produced by the generator and real ones sampled from the training data, which trains the generator to create more realistic images.

In this section, we present the results from the analyses of patterns of pixel intensity changes with changing focus and aperture size as capture parameters as well as some generated images using curve-fitting and regression analysis. The generation of images using GAN is still work on progress.

4.1. Image Analysis and Visualization

First, we present the results from the analyses of patterns of pixel intensity changes with changing focus and aperture size of a DSLR camera to have a better understanding of the datasets. FIGURE 4 presents some of the results from the analysis. FIGURE 4 (a) shows one sample image from one of the datasets which contain twenty-three photos or frames taken by changing the aperture size of the camera as explain in FIGURE 3, and FIGURE 4(b) shows the changes of the grayscale pixel intensity values in each successive frames. The horizontal axis represents the sequence number of the frames, while the vertical axis presents the sequential number of the pixels. Each line shows the gradual intensity change of a single-pixel. FIGURE 4(c) is the visual representation of the variance of each pixel with respect to the changes in the sequential frames, and FIGURE 4(d) is showing the changes of the grayscale intensity of only pixels with higher variance than a predefined threshold value. FIGURE 4(e) is the visual representation of the maximum gradient values for each pixel with respect to the changes in the sequential frames, and FIGURE 4(e) is showing the changes of the grayscale intensity of only pixels with higher maximum gradient than a predefined threshold value. Approaches used in FIGURE 4(c) - (f) are filtering out the pix-

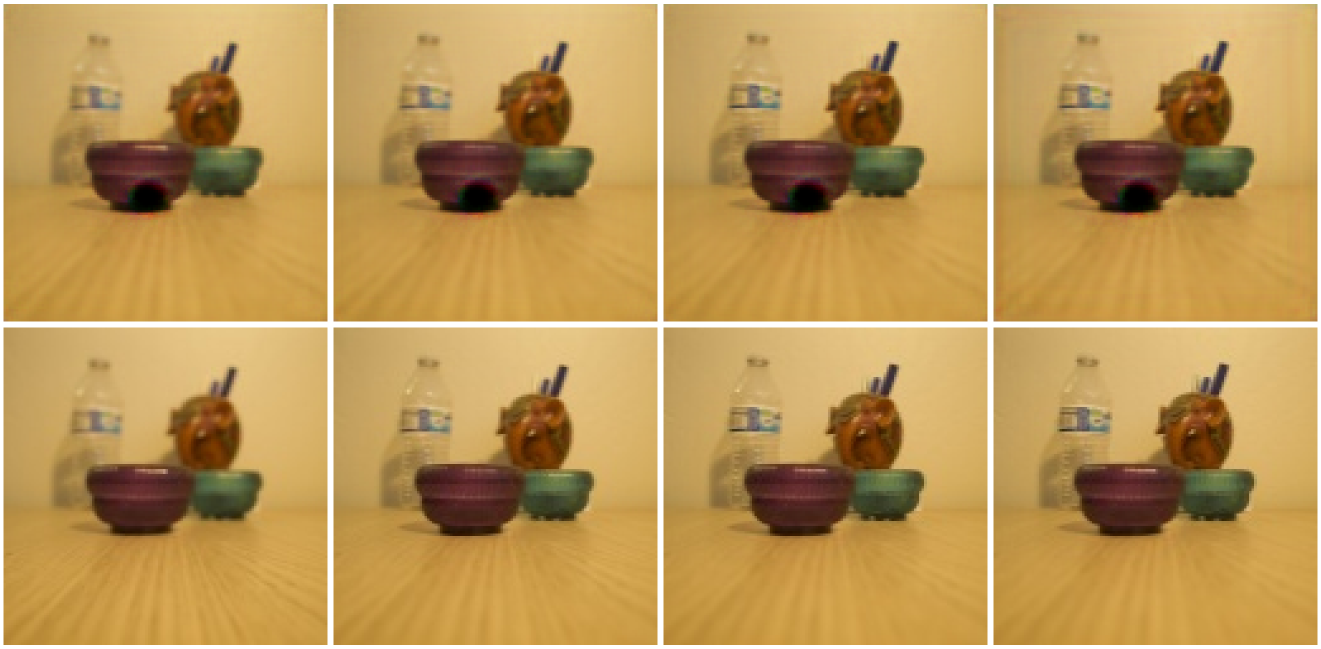


Figure 10. Results generated by GAN. left : near focus, right : far focus: top row generated images : bottom row ground truth images

els with the behavior different than the overall image. These analysis help us to understand the underlaying pattern of the image data and change of pixel behaviours with change of focus and aperture size. We also created a visualization tool for comparative visualization of change of pixel behaviours in original and generated images by clicking and selecting pixels from the images (FIGURE 8).

4.2. Results: Curve fitting based methods

For the image synthesis, our first approach was generating the images using curve fitting by linear and piecewise cubic polynomials. The results of generating images by using a linear curve fitting are presented in FIGURE 5. The left side or column 5(a) has four samples of original images, that can be considered as ground truth. The right side or column 5(b) presents the four samples of generated images respective to the images on the left side. Although the generated images mostly look natural and realistic, the first and third image has some "halo" like artifacts. Images generated using curve fitting by piecewise cubic polynomials have similar issues and some results are presented in FIGURE 6. Even though the Root Mean Square (RMS) Error for the generated images are not very high (FIGURE 7), these results prove that any straight forward curve fitting approach can not optimally generate images for all capture parameters.

4.3. GAN based method : preliminary results

We configured our GAN to generate images at four different focus levels. As shown in Figure 10, the generated images to the left are with near focus and the generated images to the right are with far focus.

5. Discussion

In our work, initially, we analyzed the datasets to discover the pattern and relationship among the changing pixels with the changing of different capture parameters for understanding the challenges of the research problem. Some results of this brief analysis are presented in FIGURE 4. After that, we generated some images by using linear and cubic curve-fitting approaches to analyze the limitations of this approach. The limitations are apparent from the results presented in FIGURE 5. For generating images using a simple machine learning approach, we used regression analysis. We are working on a GAN based approach for our final results.

6. Conclusion

In this paper we present a approach for generating all the possible images that can be taken with all possible setup of capture parameters (focus and aperture size) of a DSLR camera just by using a few photos. For that purpose, we created several datasets of images using a set of capture parameters by varying focus and aperture size. We compared the generated images from our GAN based approach with images generated using the curve fitting and regression analysis based approaches. We also discussed our work in contrast with the works. We also provided some guidelines about the extension of this work and potential future works.

References

[1] Agarwala Aseem, Dontcheva Mira, Agrawala Maneesh, Drucker Steven, Colburn Alex, Curless1 Brian, Salesin David, and Cohen Michael. Interactive digital photomontage. In *SIGGRAPH '04 ACM SIGGRAPH 2004 Papers*, pages 294–302, 2004. 2

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 3, 6

[3] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 3

[4] David E Jacobs, Jongmin Baek, and Marc Levoy. Focal stack compositing for depth of field control. *Stanford Computer Graphics Laboratory Technical Report*, 1(1):2012, 2012. 2

[5] David E Jacobs, Orazio Gallo, and Kari A Pulli. Dynamic image stacks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 144–151, 2014. 2

[6] Marc Levoy and Pat Hanrahan. Light field rendering. In *SIGGRAPH '96 Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, 1996. 3

[7] Kevin Matzen, Vivien Nguyen, Dillon Yao, You Zhang, Ren Ng, et al. Synthetic defocus and look-ahead autofocus for casual videography. *arXiv preprint arXiv:1905.06326*, 2019. 2

[8] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3

[9] Ren Ng. Fourier slice photography. In *ACM Transactions on Graphics (TOG) Volume 24 Issue 3*, pages 735–744, 2005. 3

[10] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Gaugan: semantic image synthesis with spatially adaptive normalization. In *ACM SIGGRAPH 2019 Real-Time Live!*, page 2. ACM, 2019. 3

[11] Parikshit Sakurikar, Ishit Mehta, Vineeth N. Balasubramanian, and P. J. Narayanan. Refocusgan: Scene refocusing using a single image. In *The European Conference on Computer Vision (ECCV)*, September 2018. 3

[12] W. Hasinoff Samuel and N. Kutulakos Kiriakos. Light-efficient photography. In *ECCV 2008: Computer Vision – ECCV 2008*, pages 45–59, 2008. 2

[13] Christian J. Schuler, Harold Christopher Burger, Stefan Harmeling, and Bernhard Scholkopf. A machine learning approach for non-blind image deconvolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013. 3

[14] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 3

[15] Wenqi Xian, Patsorn Sangkloy, Varun Agrawal, Amit Raj, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. Texturegan: Controlling deep image synthesis with texture patches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8456–8465, 2018. 3