JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

SYNTHETIC REFOCUSING OF AN IMAGE USING MONOCULAR DEPTH ESTIMATION

[1]Sarwesh Nandeshwar, [2]Mohit Shukla, [3]Rahul Singh and [4]Kirti Rajadnya
[1,2,3]Scholar, [4]Professor
Department of Information Technology,
SSJCOE, Dombivli, India

Abstract: This model uses deep learning and computational photography techniques to create synthetic refocusing of images. It lets you control the depth of field and focus on specific objects in a range of applications, such as computer vision, photography, and microscopy. The model is flexible and adaptable, thanks to its ability to estimate depth maps by analyzing spatial frequencies and depth cues. This results in highly precise depth maps that can be used to create multiple versions of an image with varying levels of blur. This allows you to have more control over the refocusing process, giving you the ability to highlight specific regions or objects within an image. Synthetic refocusing is a useful tool for both creative and practical aspects of image focus control, and this model provides valuable insights for researchers and practitioners alike.

Index Terms - Synthetic Refocusing, Depth Map, Computer Vision, Deep Learning, Depth of Field.

1. INTRODUCTION

The pursuit of manipulating post-capture image focus has driven advancements in photography and computer vision. Synthetic Refocusing emerges as a potent tool, offering control over focus and background elements, enabling artistic expression and practical applications across diverse fields. Traditionally, refocusing relied on capturing multiple images at various focal distances, demanding specialized hardware and substantial computational resources. However, recent strides in monocular depth estimation techniques have revolutionized synthetic refocusing, making it more accessible by facilitating the process.

This approach to synthetic refocusing involves monocular depth estimation, utilizing computer vision and machine learning algorithms to predict object depth from a single image. This enables the creation of depth maps, facilitating post-capture adjustments to the focus plane and generation of images with diverse depths of field. The process employs advanced deep learning algorithms, training neural networks on extensive datasets pairing images with corresponding depth information. Stereo Depth Estimation is a technique in computer vision that aims to determine the depth information of a scene by analysing two or more images captured from slightly different viewpoints. This technique works similarly to human vision, where our brain processes the differences in the images seen by each eye to perceive depth and 3D information. We plan to replicate the method used in Stereo Depth Estimation and implement it using Monocular Depth Estimation. This involves using two copies of a single image with slight variations, different viewpoints, and changes in angle by a small degree and processing them in a neural network to obtain depth information. It has the potential to be effective. Left-right consistency becomes climactic in monocular depth estimation, ensuring alignment accuracy in stereo-vision scenarios, including robotics and autonomous navigation. This character verification enhances confidence in the accuracy of depth maps produced during the monocular depth estimation process.

This method attempts to directly predict the depth of each pixel in an image using models that have been already trained on large collections of data. The fundamental problem in machine perception, is to understand the shape of the scene form a single image. In the past, adjusting camera settings like aperture and lens position was the only way to refocus images during capture. However, with the recent advancements in monocular depth estimation, we now have synthetic refocusing. This technique allows for post-capture adjustment of focus and depth of field. It is versatile and creative, making it an excellent tool for photographers, image editors, and professionals in various fields, including computer vision. Synthetic refocusing is helpful in emphasizing specific elements within an image, enhancing storytelling, and artistic potential. It can also be used in medicine, virtual reality, and robotics, where quick adjustments are essential. Overall, synthetic refocusing is a valuable tool for enhancing images and facilitating creative expression across different domains.

When training models to generate depth information, it is important to use a loss function that accurately assesses performance. One such function that can be quite helpful in stereo vision tasks is the Left-Right Disparity Consistency Loss. By ensuring that estimated depth information from both left and right views is aligned properly, this function can greatly enhance the reliability and precision of the generated data. Synthetic refocusing proves versatile, finding applications in photography, filmmaking, medicine, virtual reality, 3D modelling, object recognition, autonomous vehicles, forensic analysis, microscopy, surveillance, and education.

The ability to adjust focus post-capture holds practical significance, enabling creative effects, correcting focus errors, enhancing medical diagnostics, seamlessly integrating virtual and real worlds, and contributing to various scientific and technological domains.

1.1: OBJECTIVES

1) Post-Capture Adjustment: Develop a model that allows users to modify the focus and depth of field of an image after it has been captured, providing flexibility and creative control in image editing.
2) Enhanced Depth Perception: Improve depth perception in images by accurately adjusting focus and depth of field, making it easier to discern the relative distances of objects in the scene and enhancing the overall visual experience.
3) Error Correction: Provide a mechanism to correct focus errors or missed opportunities during image capture, allowing users to refine the focus and quality of the final image post-capture.
4) Correcting Blur or Defocus: Develop algorithms to correct blur or defocus in images that were initially captured with incorrect focus, ensuring that the subject becomes clear and well-defined after refocusing.
5) Emphasizing a Subject: Enable users to highlight a particular subject or element in the image by making it the focal point, allowing for creative expression and emphasis on key elements of the scene.

2. LITERATURE REVIEW

This paper [1] proposes a novel end-to-end unsupervised monocular depth estimation network that leverages the Multi-Orientation Epipolar Geometry of Light Field. The system introduces three unsupervised loss functions based on the inherent geometry constraints and depth cues of the light field. The model can predict the depth of the light field without any ground-truth information and has been rigorously tested on both synthetic and real-world datasets. The method offers three key advantages: it can learn the light field depth without requiring any ground-truth information, it can be extended seamlessly to numerous light field applications, and it can be integrated into more complex and larger light field applications.

The model in this paper [2] suggests using computer-generated images from video games to estimate the depth of objects in a picture taken with a single camera. This method can help create a 3D image from a 2D photo. The computer-generated images are created in a virtual environment that is designed for video games. The model takes advantage of style transfer and adversarial training to predict pixel-perfect depth from a single real-world colour image based on training over a large corpus of synthetic environment data. However, the model cannot perform well on real-world data as the domain distributions to which these two sets of data belong are widely different. To overcome this issue, the authors propose the use of a GAN-based style transfer approach to adapt their real-world data to fit into the distribution approximated by the generator in the depth estimation mode.

This paper [3] presents a new method to estimate the depth of an image using a deep neural network. Unlike traditional methods, this approach does not require expensive ground truth depth data. Instead, it leverages binocular stereo data to obtain the depth estimates. During training, a novel loss function is utilized to ensure that the predicted depth maps from each camera view are consistent with each other, leading to improved accuracy. The results of this method are better than traditional methods that rely on ground truth data. Furthermore, the model can generalize well to new datasets and produce visually plausible depth maps. In the future, the model can be extended to work with videos by adding temporal consistency to the depth estimates. Additionally, sparse input could be explored as an alternative training signal.

The proposed system in this paper [4] uses graded blurring to refocus on specific parts of an original image. To do this, it generates a depth map of the image using Monocular Depth Estimation (MDE) with Machine Learning. The system then stitches together different blurred images based on the distance of a point from the depth of focus. This process is designed to dynamically refocus an image, improve the details of the refocused part, and require fewer resources. The implementation involves model training, image preprocessing, depth map generation, normalization of depth values, and blurring of the image.

This paper [5] proposes a fully convolutional architecture to estimate the depth map of a scene given a single RGB image. The architecture includes residual learning and efficiently learns feature map up-sampling within the network to improve the output resolution. The reverse Huber loss is introduced for optimization, which is particularly suited for the task at hand and driven by the value distributions commonly present in in-depth maps. The model is composed of a single architecture that is trained end-to-end and runs in real-time on images or videos. The proposed network is fully convolutional, comprising up-projection layers that allow for training much deeper configurations, while greatly reducing the number of parameters to be learned and the number of training samples required.

This paper [6] discusses a method that aims to address the challenges of monocular depth estimation. The approach proposed in the paper involves modelling the 3D motion of individual objects to predict depth information without relying on depth sensors. The method leverages the inherent structure in monocular videos for unsupervised learning, using computer vision techniques to estimate depth from single images or video sequences without the need for specialized depth-sensing hardware or manual labelling of depth data.

3. METHODOLOGY

The process of refocusing an image begins with an input image. Preprocessing of the input image is performed to enhance its quality and consistency. Then, a Convolutional Neural Network (CNN) is used to analyze the pre-processed image and generate a corresponding depth map. The depth map provides valuable information about the spatial dimensions and distances within the scene depicted in the input image. Based on this depth data, a new image is synthesized, which incorporates adjustments to the focus plane and depth of field as desired. Finally, graded blurring techniques are applied to the generated image to achieve the desired refocusing effect. These techniques selectively adjust the sharpness of different regions based on their distance from the focal plane. This comprehensive process enables the transformation of a single input image into a refocused output, which provides enhanced visual clarity and depth perception.

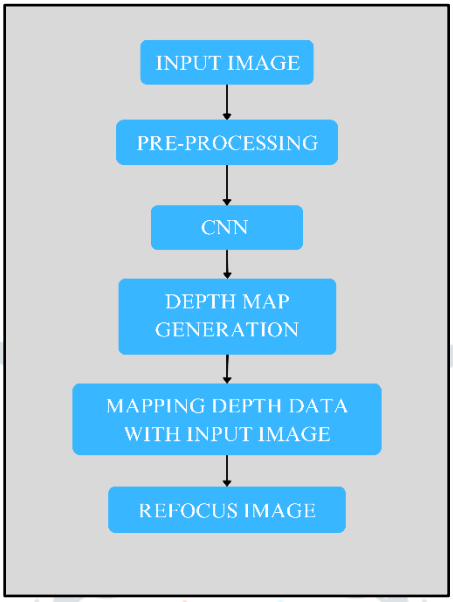

Figure 3.1: Flow Chart

1.2: Proposed Architecture

The process begins with an input image, which is fed into a Convolutional Neural Network (CNN) for analysis. The CNN extracts feature and generates a disparity map representing the depth information of the scene. This disparity map undergoes matrix multiplication to refine the depth values. The resulting depth map is then used to warp the input image, adjusting the focus and depth of field based on the depth information. Finally, the warped image, representing the refocused output, is generated. This is shown in the Figure 3.1.1:

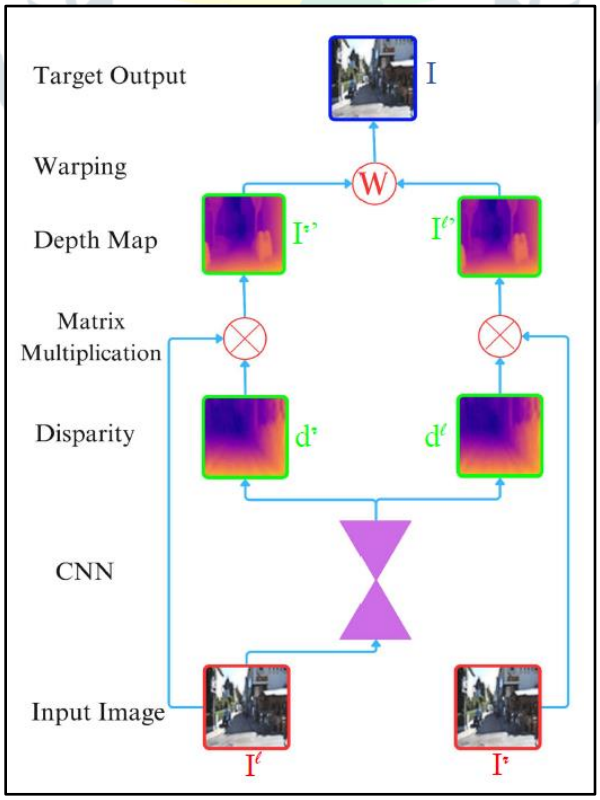

Figure 3.1.1: Network Architecture of the System

1.3: Algorithm

Step 1: Input –

- Provide the input image (input_image).
- Obtain the depth map of the input image (depth_map).
- Select the blur method (blur_method).

Step 2: Generate Depth Map –

- The system runs the depth.py script, which uses the MiDaS DPT model.
- This model is based on a transformer architecture and takes a single RGB image as input.
- The depth map represents the estimated distance of objects from the camera and is crucial for subsequent steps.
- During inference, the model processes the input image to generate the depth map.
- Once generated, the depth map is ready for use in the refocusing process.

Step 3: Defocus Image –

- Initialize a DefocuserObject with the input image path (input_image) and the selected blur method (blur_method).
- Load the depth map (depth_map) corresponding to the input image.
- Generate blurred versions of the input image using the selected blur method.
- Normalize the depth map around the point of focus.
- Define a mouse callback function to set the point of focus based on user interaction with the depth map.
- When the user clicks on the depth map, set the point of focus and perform defocusing around that point.
- Apply masking to each blurred image based on the normalized depth map.
- Combine the masked images to generate the final defocused image.

Step 4: Output –

- The system produces the final defocused image.
- This image shows the simulated focus adjustment based on the depth map.
- Users can view or save the defocused image for further analysis.
- It provides insights into the depth perception within the scene captured by the input image.

4. RESULTS

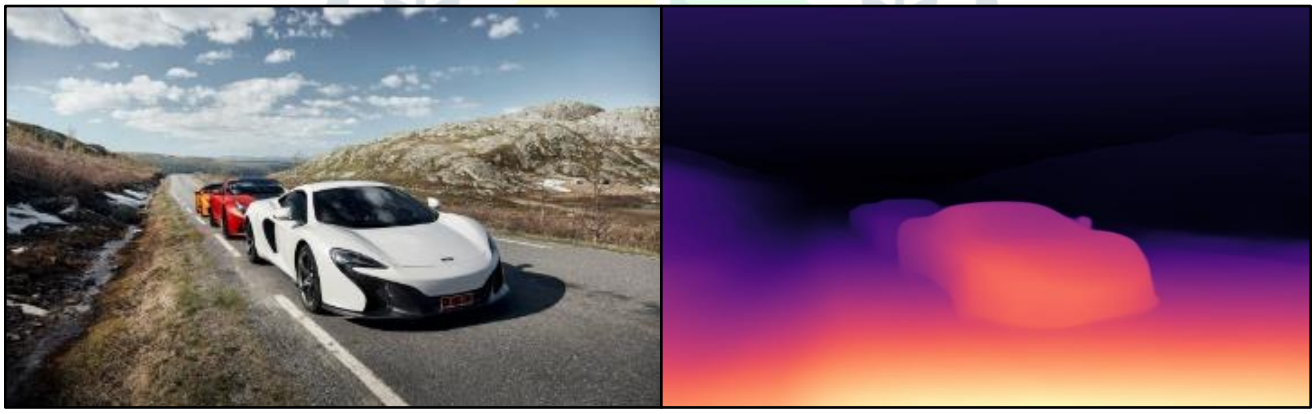

Figure 4.1: Input Image of a Car and it's Depth Image

Figure 4.2: Input Image of a Village and it's Depth Image

Figure 4.3: Input Image of an Alley and it's Depth Image

Figure 4.1, 4.2, and 4.3 represents depiction of Input Images and their Depth Mapped Images or Depth Maps. Based on the distance of objects in the image from the focal point, objects are represented with different colors using warm and cool colors. In these figures, the Depth Map is displayed in the MAGMA type of depth representation.

Figure 4.4: Input Image

Figure 4.5: Refocused Image

Figure 4.6: Input Image and it's Refocused Image

Figure 4.4, 4.5, and 4.6 represents the Input Image and its Refocused Image after processing in the system. The first image (Figure 4.4) is the original image which is used as the input to the system. The second image (Figure 4.5) is the refocused image of

the input image. In the second image the right-most corner is refocused and according to the focal point, i.e., the distance from the focus point, rest of the image is slightly blurred or defocused using graded blurring technique.

5. CONCLUSION

The project focuses on enhancing image refocusing capabilities through monocular depth estimation. Leveraging machine learning algorithms and neural network, the project achieves precise depth prediction from single input images, enabling control over focus planes and depth of field post-capture. Notably, the project contributes to advancing imaging techniques' accessibility and usability by streamlining refocusing process and eliminating the need for specialized hardware.

Looking forward, the insights gained from the project pave the way for further developments in monocular depth estimation and image refocusing, promising new opportunities in visual storytelling, scientific analysis, and immersive experiences.

6. FUTURE WORK

As part of future work, we plan to conduct comprehensive evaluations and comparisons to better understand the performance of our project relative to other models and methodologies. This assessment will involve benchmarking against existing approaches to identify areas for improvement and refine our algorithms. Additionally, we intend to explore the utilization of alternative depth map generation algorithms, such as stereo depth estimation, for similar applications. By integrating multiple approaches, we aim to enhance the project's capabilities and robustness, leveraging the strengths of each method to achieve more accurate and reliable depth predictions across diverse scenarios. Collaborating with researchers and practitioners in the field will be essential for gaining insights and driving innovation in image refocusing and depth estimation.

NOMENCLATURE

Sr. No.	Abbreviation	Expansion
1.	GAN	Generative Adversarial Network
2.	MDE	Monocular Depth Estimation
3.	RGB	RED, GREEN, And Blue
4.	CNN	Convolutional Neural Network
5.	DPT	Depth Prediction Transformers

Figure 1: Nomenclature

ACKNOWLEDGMENTS

We would like to express our gratitude to our college Principal, Dr. Pramod R. Rodge, for providing us with lab facilities and allowing us to pursue our project. We also extend our heartfelt thanks to our H. O. D., Dr. Savita S. Sangam, who provided us with the necessary computer facilities in our laboratory and played a significant role in making our project a success. Without their cooperation, our project would have come to a standstill. We would also like to thank our Project Guide, Prof Kirti Rajadnya, for her unwavering support throughout the project. Her expert guidance, kind advice and timely motivation have been invaluable in helping us develop our project, "Synthetic Refocusing of an Image Using Monocular Depth Estimation".

We would like to acknowledge the contributions of our colleagues who supported us directly or indirectly during the project. In conclusion, we sincerely thank everyone who has contributed to our project and helped us achieve success.

REFERENCES

[1] Wenhui Zhou, Enci Zhou, Gaomin Liu, Lili Lin, and Andrew Lumsdaine, “Unsupervised Monocular Depth Estimation From Light Field Image”, IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 29, 2020.

[2] Rui Guo, Babajide Ayinde, Hao Sun, Haritha Muralidharan and Kentaro Oguchi, “Monocular Depth Estimation Using Synthetic Images With Shadow Removal*”, IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, NZ, October 27-30, 2019.

[3] Cl´ement Godard, OisinMac Aodha, Gabriel J. Brostow, “Unsupervised Monocular Depth Estimation with Left-Right Consistency”, University College London, 12 April, 2017.

[4] Keerthi Balathoti, Gatikoppu Sahithi, Himanshu Indugamelli, Priyanka Dundi, “Graded blurring for image refocusing using Monocular Depth Estimation”, February 2023.

[5] Iro Laina, Christian Rupprecht, Vasileios Belagiannis, Federico Tombari, Nassir Navab, “Deeper Depth Prediction with Fully Convolutional Residual Networks”, IEEE International Conference on 3D Vision (3DV) 2016

[6] Vincent Casser∗1 Soeren Pirk Reza Mahjourian2 Anelia Angelova, “Depth Prediction Without the Sensors: Leveraging Structure for Unsupervised Learning from Monocular Videos”, 15 November 2018

[7] http://visual.cs.ucl.ac.uk/pubs/monoDepth/

[8] https://youtu.be/go3H2gU-Zck?si=-CakCmyTjjn3b8n0

[9] https://github.com/batman-nair/project-defude

[10] https://github.com/mrharicot/monodepth?tab=readme-ov-file

[11] https://huggingface.co/spaces/LiheYoung/Depth-Anything/tree/main/checkpoints

[12] https://github.com/heyoeyo/muggled_dpt