ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

# 2D to 3D Image Conversion Using Machine Learning Approach

Rameshwar Dhondge[1], Shreya Jadhav[2], Atul Gunjal[3], Pranjali Joshi[4]

B.E. Student, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India[1,2,3]

Assistant Professor, Department of Computer Engineering, Pune Institute of Computer Technology, Pune, Maharashtra, India[4]

**ABSTRACT**: There is a lot of advancement in the virtual reality (VR) market, due to this there is interest in creating 3D content. Current techniques which are used for creating 3D content is very tedious and costly. For shooting 3D movies whole different set is required also the stereo cameras are required which are very costly. Generally, these cameras are not affordable by everyone. In this paper, our aim is to convert the 2D image/movie to 3D image/movie using deep learning techniques. Our approach is to train the Deep Convolutional Network (CNN) on stereo pairs of images which are extracted from 3D movies. The Deep CNN outputs the depth map using a single image as input. This depth map can be combined with the input image to produce the corresponding stereo image of the input image.

**KEYWORDS**: Machine Learning, Virtual reality, Convolutional Network, Deep Learning, Stereoscopy, Depth map.

## I. INTRODUCTION

Stereoscopy is the process of producing the illusion of depth in the two-dimensional image by means of stereopsis for binocular vision. For normal 2D images, we can add additional dimension 'DEPTH' to perceive them as 3D. Stereoscopy process creates stereo-pair of a 2D image. Stereo-pair contains two images one of the views is intended for the left eye and the other for the right eye.

To create an illusion of depth we need to generate the depth map of the corresponding image. Depth map gives information about per pixel depth of different objects placed in the image.The most important and difficult problem in converting 2D to 3D is how to produce or estimate the depth map using a single view of an image. Traditionally the depth map can be calculated by using the Stereo camera, Laser triangulation, etc techniques. For 3D movies, depth map calculation process is the human intervention process which completely relying on “depth artists”.

(a) (b)

Fig. 1. Depth map (brighter pixels are closer to camera) (a) Original image (b) Corresponding depth map.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

In an automatic approach of creating depth maps, the process consists of two steps:

1. Calculating depth map from input view.
2. Combine depth and input view to generate missing view of stereo pair.

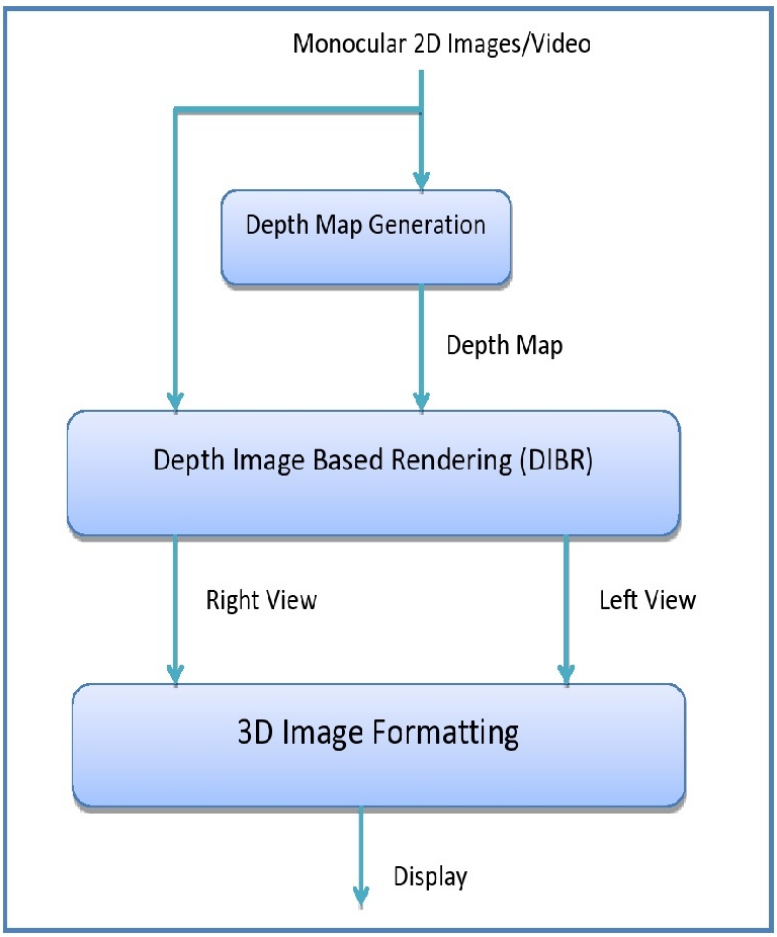


Fig. 2. Overview of system

Paper is organized as follows. Section II describes related work for converting 2D to 3D, section III describes data collection and pre-processing. Section IV represents proposed system. Section V presents implementation of system Section VI presents experimental results showing results of images tested. Finally, Section VII presents conclusion.

## II. RELATED WORK

The first model proposed for 2D to 3D was Deep3d [2], proposed by Xie et al. [2016]. The core idea of Deep3d was to use a neural network for prediction of disparity map and using disparity map we can generate right view of image from left view. However, Deep3d image formation model is not fully differentiable. Another drawback of Deep3d was

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

That the size of image is fixed. They [2] address this by predicting a distribution over a range of disparity values at each pixel, instead of predicting a single value. The right image is then obtained by computing the expected value at each pixel in the right image over the range of disparity values.

In [1] the authors proposed a fully convolutional deep neural network loosely inspired by the supervised DispNet architecture of Mayer et al. [10]. They treat depth estimation problem as an image reconstruction problem their architecture can solve for the disparity field without requiring ground truth depth hence learning is unsupervised. They are using a left-right consistency check to improve the quality of synthesized depth images.

Given a single image I at test time, their goal is to learn a function f that can predict the per-pixel scene depth, $\hat{d}=f(I)$. Specifically, at training time, they have access to two images $I^l$ and $I^r$ , corresponding to the left and right color images from a calibrated stereo pair, captured at the same moment in time. They attempt to find the dense correspondence field $d^r$ that, when applied to the left image, would enable to reconstruct the right image.

## III. DATASET COLLECTION

Our Deep Convolutional Neural Network model is designed to train on stereo pairs. To generate training data, we could fully utilize the current existing 3D videos. The dataset required for training is the left and right images respectively. To achieve this goal, we acquired three 2-hours full HD movies Titanic, Need for Speed and Jumanji with the resolution of 1920 by 1080. We collected a total of 33700 frames from all 3D movies and layout left and right frames next to each other. During the data processing, we re-sized cut videos to lower resolution to 512 by 256 to process them fast to generate stereo pairs. For pre-processing purpose, we have used OpenCV and Python. We also removed the video parts which would influence the training like the frames with casting information and VFX animation.

## IV. PROPOSED SYSTEM

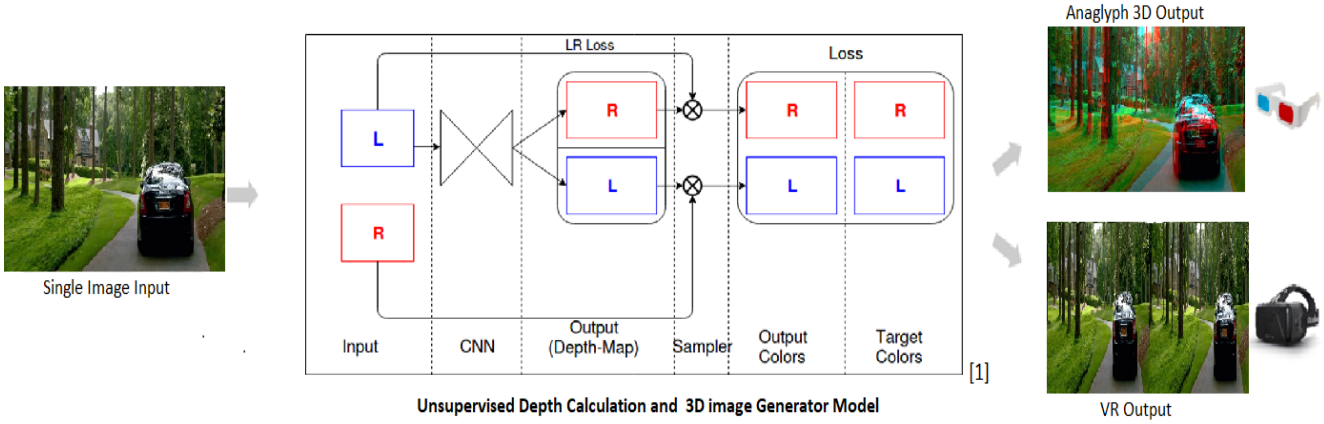


Fig. 3. Proposed system

For depth-map calculation from a single image, we are using Deep Convolutional network (CNN) proposed in [1]. The goal of CNN [1] is to produce a depth map from a single input image instead of giving stereo pairs as input. In this unsupervised approach, we are giving stereo images as input, unlike previous supervised approaches where the input is stereo images along with actual disparity. Once we get a depth map it's easy to find the corresponding view of a stereo pair. At a high level, network estimates depth by inferring the disparities that warp the left image to match the right one. The key to this method is using only the left input image obtain a depth map by enforcing them to be consistent with

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

stereo pairs using backward mapping and bilinear sampling. For generating left depth map we need to train CNN correspondingly.

CNN has two main parts - an encoder (from cnv1 to cnv7b) and decoder (from upcnv7), The decoder uses skip connections from the encoder's activation blocks, enabling it to resolve higher resolution details. We output disparity predictions at four different scales (disp4 to disp1), which double in spatial resolution at each of the subsequent scales. Even though it only takes a single image as input, the network predicts two disparity maps at each output scale - left-to-right and right-to-left.

Disparity map is scalar value per pixel which is generated from input image is used to generate right stereo view from left stereo view.In this for each pixel of coordinates (x, y) in the left stereo view is copied to the pixel on the right stereo view at the coordinates (x - d, y) where d is the value of the disparity map at the coordinates (x, y).hence we get the generated right stereo view which is shift of left stereo view and the shift depends on the value we get from disparity map.

Once we get two views our system is converting 2D to 3D in either anaglyphic form for red-cyan 3D specs or side-by-side form for VR.

## V. IMPLEMENTATION

The VGG-19 network which is implemented in Tensorflow having 19 layers, and takes about 36 hours for train using GeForce GTX 1080 GPU on a dataset of 33 thousand images. The inference is fast and takes more than 14 frames per second, for a 512×256 image, including transfer times to and from the GPU. For performing the image operations OpenCV is used. Along with OpenCV, Python libraries like Numpy, Scipy, ImageIo and PIL libraries are used. The Web-based UI is developed using Django MVT framework for accepting input and displaying output.

## VI. RESULTS

Following Figures shows some good results of our system Figs. 4 (a) shows the original left image. (b) Is the corresponding generated depth-map. (c) shows the generated right view. (d) shows the corresponding anaglyph image.

(a) (b)

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 3, March 2019

(c) (d)

Following Figures shows bad results of our system Figs. 5 (a) shows the original left image. (b) Is the corresponding generated depth-map. (c) shows the generated right view. (d) shows the corresponding anaglyph image.

(a) (b)

(c) (d)

## VII. CONCLUSION

In this paper, we proposed a 2D to 3D image conversion techniqueusing unsupervised machine learning approach. Our approach toimplementing the learning model is to generate disparity maps and thenpredict the right image from left. The results show that our methodcan generate goodresults and can handle the video input also. Alongwith some fine results from our implementation, we have found somepoor results too. When an image contains blurry part then it's hard topredict the disparity map. Also, sometimes predicted disparity map isdifferent from the actual disparity map because of predicted pixelsmay not exist in the actual image. To overcome the poor results, morecomplex methods are required which are the future plans of thisproject.

ISSN (Online): 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

*(A High Impact Factor, Monthly, Peer Reviewed Journal)*

*Visit: www.ijirset.com*

**Vol. 8, Issue 3, March 2019**

## REFERENCES

[1] Clment Godard Oisin Mac Aodha Gabriel J. Brostow,Unsupervised Monocular Depth Estimation with Left-Right Consistency University College London.CVPR 2017

[2] JunyuanXie, Ross Girshick, Ali Farhadi jxie@cs.washington.edu, ross.girshick@gmail.com, ali@cs.washington.edu ,Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks University of Washington.

[3] Zeal Prof. Lynette Dmello,Ganatra,RiddhiChavda Conversion of 2D Images to 3D Using Data Mining Algorithm International Journal of Innovations Advancement in Computer Science IJIACS ISSN 2347 8616 Volume 3, Issue 7 September 2014

[4] Nidhi Chahal , Meghna Pippal and Santanu Chaudhury Automated Conversion of 2D to 3D Image using Manifold Learning Department of Electrical Engineering Indian Institute of Technology, Delhi, India 110016

[5] Jos L. Herrera, Carlos R. del-Blanco and Narciso GarcaA Novel 2D to 3D Video Conversion System Based on a Machine Learning Approach 2015 IEEE

[6] Van Sinh NGUYEN* Manh Ha TRAN* Hoang Minh Quang VU,A Research on 3D Model Construction from 2D DICOM 2016 IEEE DOI 10.1109/ACOMP.2016.10

[7] Dineesh Mohan, Dr. A. Ranjith Ram Department of Electronics Communication Engineering Government College of Engineering Kannur,A Review on Depth Estimation for Computer Vision Applications,ISO 9001:2008 Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 4, Issue 11, May 2015

[8] Yasir Salih Aamir S. Malik, Zazilah May Dept. of Electrical Electronic Engineering UinversitiTeknologi PETRONAS Perak, Malaysia Dept. of Electrical Electronic Engineering,Depth Estimation Using Monocular Cues from Single Image 2011 IEEE

[9] Quan Wei Jiang Shan School of Computer Science and Technology Changchun University of Science and Technology, Zhang Chao, Zhang Zhenpu School of Computer Science and Technology Changchun University of Science and Technology,A real-time 2D to 3D video conversion algorithm based on image shear transformation 2015 IEEE DOI 10.1109/IMCCC.2015.362

[10] N. Mayer, E. Ilg, P. Hausser, P. Fischer, D. Cremers, A. Dosovit- ¨ skiy, and T. Brox. A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation. In CVPR, 2016. 2, 3, 4, 5