iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 Issue: VI Month of publication: June 2014

DOI:

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

A Global Nearest-Neighbour Depth Learning-based, automatic 2D-to-3D image conversion

Ms.Manasa.G.Kambi1, Prof. Nagabhushan Katte

1M.Tech(DE),Dept of ECE
1,2 Ballari Institute Of Technology And Management
Bellary,Karnataka,India

Abstract- The proposed work is to present a new method based on the radically different approach of learning the 2D-to-3D conversion from examples. It is based on globally estimating the entire depth map of a query image directly from a repository of 3D images (image+depth pairs or stereopairs) using a nearest-neighbor regression type idea.
Keywords: 3D images, Stereopairs, Image conversion, Nearest neighbor classification, Cross-bilateral filtering.

I. INTRODUCTION

The presence of 3D-capable hardware , such as TVs, Blu-Ray players, smart phones, and gaming consoles, are not yet matched by 3D content production. Despite of significant growth, 3D movies are still an exception rather than a rule, and 3D broadcasting is still minuscule compared to 2D broadcasting. In order to bridge the gap between 3D hardware and 3D content availability, many 2D to 3D conversions methods have been proposed.

A typical 2D-to-3D conversion process consists of two steps: depth estimation for a given 2D image and depth based rendering of a new image in order to form a stereo pair. While the rendering step is well understood and algorithms exist that produce good image quality, the challenge is in estimating depth from a single image (video).

There are two basic approaches for 2D to 3D conversion: One involving human operators known as semi-automatic methods. In these a skilled operator assigns depth to various parts of image Depending on sparse depth assignment, computer algorithm estimates dense depth over entire image sequence. Human operator assigns depth to various location in image to exactly define delineation of objects following depth assignment to the delineated regions.

Second is an automatic method, no human operator intervention. A computer algorithm automatically estimates the depth for an image. In these several methods have been developed which estimates shape from shading, structure from motion and depth from defocus. These methods work for restricted scenarios but not for all arbitrary scenes.

www.ijraset.com Vol. 2 Issue VI, June 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

II. STATE OF ART

Two types of 2D to 3D conversion methods: Semi-automatic methods involve human operator intervention and automatic methods which require no
operators.

A. Semi-automatic methods

This is one of the most successful approach for 2D-to-3D conversion. This method require a significant operator intervention in the conversion process, fitting delineating objects in individual frames, placing them at suitable depths, and correcting errors after final rendering. These methods are adopted by various companies such as Imax Corp., Digital Domain Productions Inc etc. Many films have been converted to 3D using this approach.

Methods involving human operators have been most successful but also time-consuming and costly. In order to lower the cost and speed up the conversion. Guttmann *et al.* [1] have proposed a dense depth recovery *via* diffusion from sparse depth assigned by the operator.

In the first step, the operator assigns relative depth to image patches in some frames by scribbling.

In the second step, a combination of depth diffusion, that accounts for local image saliency and local motion, and depth classification is applied.

In the final step, disparity is computed from the depth field and two novel views are generated by applying half of the disparity amplitude.

The focus of the method proposed by Agnot *et al.* [2] is the application of cross-bilateral filtering to an initial depth map.

Phan *et al.* [3] propose a simplified and more efficient version of the Guttmann *et al.* [1] method using scale-space random walks that they solve with the help of graph cuts flow, then applying structure-from-motion estimation and finally extracting moving object boundaries.

The operator role is to correct errors in the automatically computed depth of moving objects and assign depth in undefined areas

B. Automatic methods

The main step in 2D to 3D conversion is depth estimation from a single 2D image. Other methods, like multi view stereo, which attempt to recover depth by estimating scene geometry from multiple images not taken simultaneously. A moving camera permits structure-from motion estimation [4] while a fixed camera with varying focal length permits depth-from-defocus estimation [5].

Recently, machine-learning-inspired techniques employing image parsing have been used to estimate the depth map of a single monocular image [6]. Such methods have the potential to automatically generate depth maps, but currently work only on few types of images using carefully-selected training data.

In the quest to develop data-driven approaches to 2D-to-3D conversion we have also been inspired by the recent trend to use large image databases for various computer vision tasks, such as object recognition [7] and image saliency detection [8]. In our first attempt, we developed a method that fuses SIFT-aligned depth maps selected from a large 3D database, this approach proved to be computationally demanding [9]. Subsequently, we skipped the costly SIFT-based depth alignment and used a different metric (based on histogram of gradients) for selecting

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

most similar depth fields from a database. As a result there was no significant quality degradation but a significant reduction of the computational complexity [10]. Very recently, Karsch *et al.* [11] have proposed a depth extraction method based on SIFT warping that exceptionally follows our initial, unnecessarily complex, approach to depth extraction [9].

III. THE PROPOSED SYSTEM

2D-TO-3D CONVERSION BASED ON GLOBAL NEAREST-NEIGHBOR DEPTH LEARNING

This approach is based on observation and assumption.

The key observation is that among millions of 3D images which are available on-line, there likely exist many whose 3D content matches that of a 2D input

Assumption is that two images that are photo metrically similar also have similar 3D structure (depth).

Given a monocular query image *Q*, assumed to be the left image of a stereo pair that we wish to compute, we rely on the above observation and assumption to “learn” the entire depth from a repository of 3D images I and render a stereo pair in the following steps:

1. Search for representative depth fields

2. Depth fusion

3. Depth smoothing

4. Stereo rendering

search for representative depth fields: find *k* 3D images in the repository I that have most similar depth to the query image

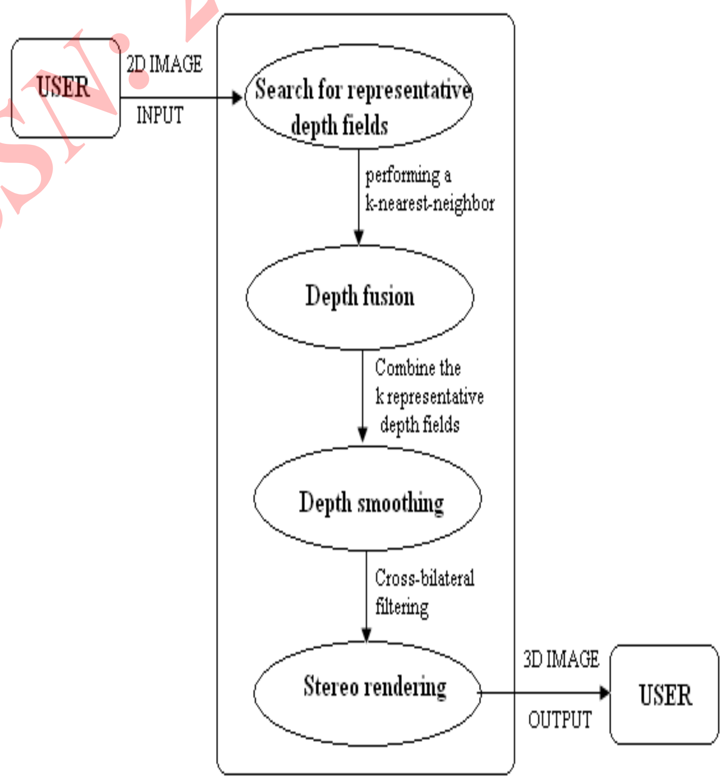
depth fusion: combine the *k* representative depth fields.

depth smoothing: process the fused depth field to remove spurious variations, while preserving depth discontinuities

stereo rendering: generate the right image of a fictitious stereo pair using the monocular query image

and the smoothed depth field followed by suitable processing of occlusions and newly-exposed areas.

Fig 1:DATA FLOW GRAPH:

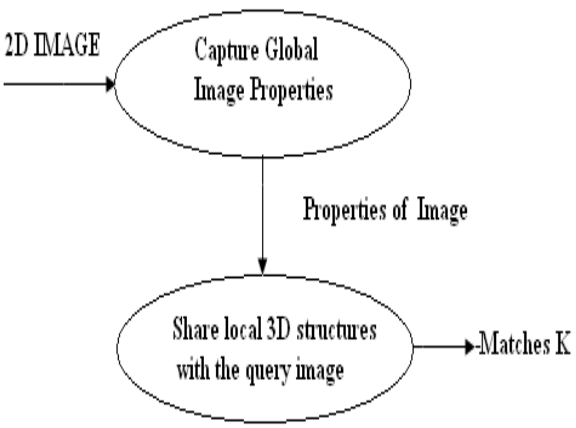


Vol. 2 Issue VI, June 2014

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Fig2: Search for representative depth fields



KNN search

There are two types of images in a large 3D image repository: one that are relevant for determining depth in a 2D query image, and second that are irrelevant.

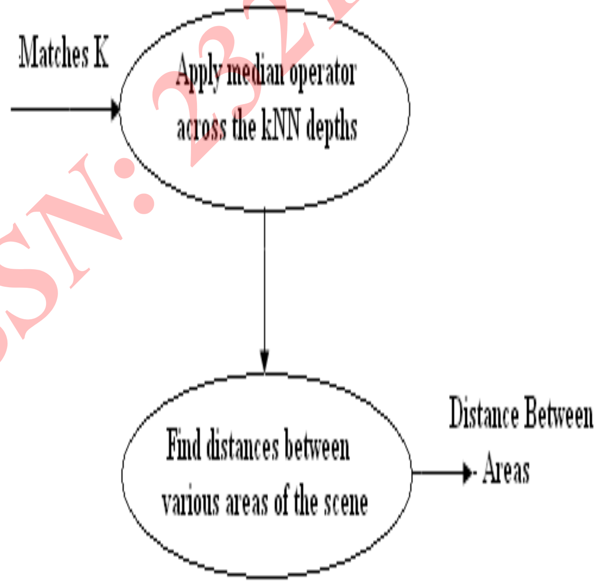
Images that are not photo metrically similar to the 2D query need to be rejected because they are not useful for estimating depth (as per our assumption).

Although we might miss some depth-relevant images, we are effectively restricting the number of irrelevant images that could potentially be more harmful to the 2D-to-3D conversion process.

The selection of a smaller subset of images provides the added practical benefit of computational tractability when the size of the repository is very large.

One method for selecting a useful subset of depth relevant images from a large repository is to select only the *k* images that are closest to the query where closeness is measured by some distance function capturing global image properties such as color, edges, textures, etc. As this distance function, we use the Euclidean norm of the difference between histograms of oriented gradients (HOGs) computed from two images. Each HOG consists of 144 real values (4×4 blocks with 9 gradient direction bins) that can be efficiently computed.

Fig3: Depth fusion

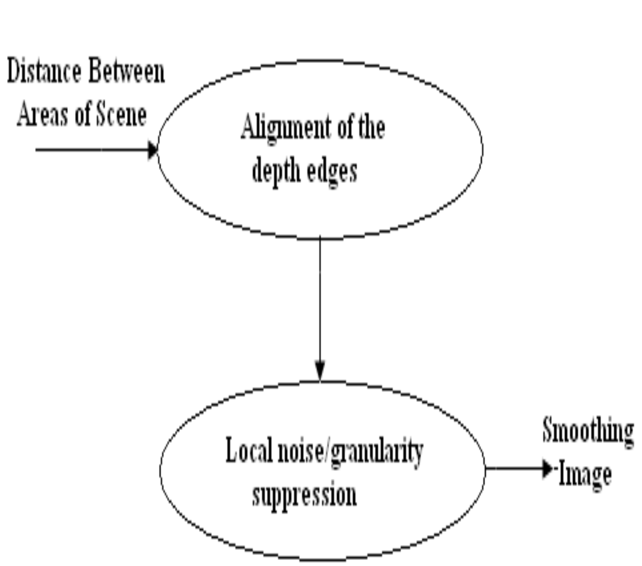


In general, none of the NN image + depth pairs (*Ii*, *di*), *i* K match the query *Q* accurately. However, the location of some objects (e.g., furniture) and parts of the background (e.g., walls) is quite consistent with those in the respective query. If a similar object (e..g, building, table) appears at a similar location in several *k*NN images, it is likely that such an object also appears in the query, and the depth field being sought should reflect this.

ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Fig4:Cross Bi-lateral filtering of depth

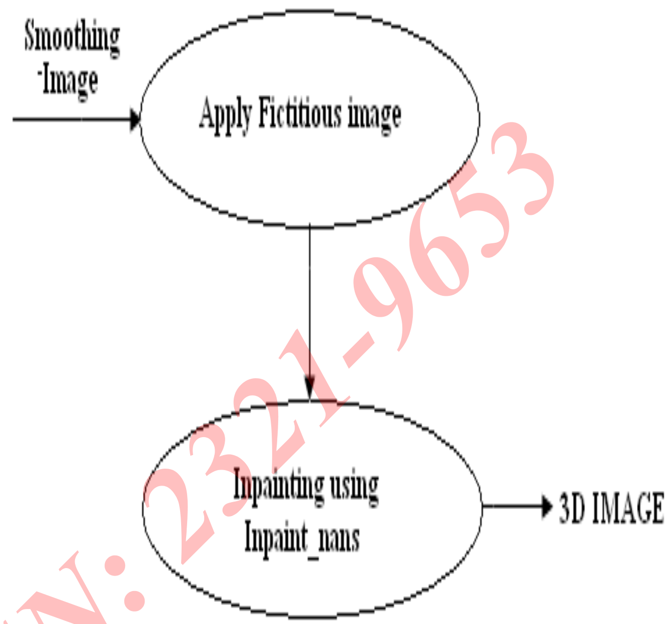


While the median-based fusion helps make depth more consistent globally, the fused depth is overly smooth and locally inconsistent with the query image due to edge misalignment between the depth fields of the *k*NNs and the query image. This, in turn, often results in the lack of edges in the fused depth where sharp object boundaries should occur and/or the lack of fused-depth smoothness where smooth depth is expected.

5. Stero rendering

In order to generate an estimate of the right image b *QR* from the monocular query *Q*, we need to compute a disparity δ from the estimated depth b*d*. Assuming that the fictitious image pair (*Q*, b*QR*) was captured by parallel cameras with baseline *B* and focal length *f* , the disparity is simply δ[*x*, *y*] = *Bf* /b*d*[**x**], where **x** = [*x*, *y*]*T* .

Fig5:Stero rendering



IV. CONCLUSION

In this paper,we discussed on globally estimating the entire depth field of a query directly from repository of image+depth pairs using nearest neighbor regression . With the continuously increasing amount of 3D data on-line and with the rapidly growing computing power in the cloud, the proposed framework seems a promising alternative to operator-assisted 2D-to-3D image and video conversion.

ACKNOWLEDGEMENT

I take this opportunity to express my profound gratitude and deep regards to my guide Prof. Nagabhushan Katte, Ballari Institute Of Technology And Management, Bellary for his exemplary guidance and encouragement throughout

REFERENCES

[1] M. Guttmann, L. Wolf, and D. Cohen-Or. Semi-automatic stereoextraction from video footage. In *Proc. IEEE Int. Conf. ComputerVision*, pages 136–142, Oct. 2009.

[2] L. Agnot, W.-J. Huang, and K.-C. Liu. A 2D to 3D video and imageconversion technique based on a bilateral filter. In *Proc. SPIE Three-Dimensional Image Processing and Applications*, volume 7526, Feb.2010.

[3] R. Phan, R. Rzeszutek, and D. Androutsos. Semi-automatic 2D to3D image conversion using scale-space random walks and a graphCuts based depth prior. In *Proc. IEEE Int. Conf. Image Processing*,

Sept. 2011.

[4] R. Szeliski and P. H. S. Torr. Geometrically constrained structurefrom motion: Points on planes. In *European Workshop on 3D Structure from Multiple Images of Large-Scale Environments*, pages

171–186, 1998.

[5] M. Subbarao and G. Surya. Depth from defocus: A spatial domain approach. *Intern. J. Comput. Vis.*, 13:271–294, 1994.

[6] A. Saxena, M. Sun, and A. Ng. Make3D: Learning 3D scenestructure from a single still image. *IEEE Trans. Pattern Anal.Machine Intell.*, 31(5):824 –840

[7] A. Torralba, R. Fergus, and W. T. Freeman. 80 million tiny images:A large data set for nonparametric object and scene recognition.

IEEE Trans. Pattern Anal. Machine Intell., 30(11):1958–1970, Nov.2008.

[8] M. Wang, J. Konrad, P. Ishwar, K. Jing, and H. Rowley. Image

saliency: From intrinsic to extrinsic context. In *Proc. IEEE Conf.Computer Vision Pattern Recognition*, pages 417–424, June 2011.

Machine Intell., 21(8):690–706

[9] J. Konrad, G. Brown, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee.

Automatic 2D-to-3D image conversion using 3D examples from the Internet. In *Proc. SPIE Stereoscopic Displays and Applications*,

[10] J. Konrad, M. Wang, and P. Ishwar. 2D-to-3D image conversion bylearning depth from examples. In *3D Cinematography Workshop*

(3DCINE'12) at CVPR'12, pages 16–22, June 2012.

[10] M. Liao, J. Gao, R. Yang, and M. Gong. Video stereolization:Combining motion analysis with user interaction. *IEEE Trans.*

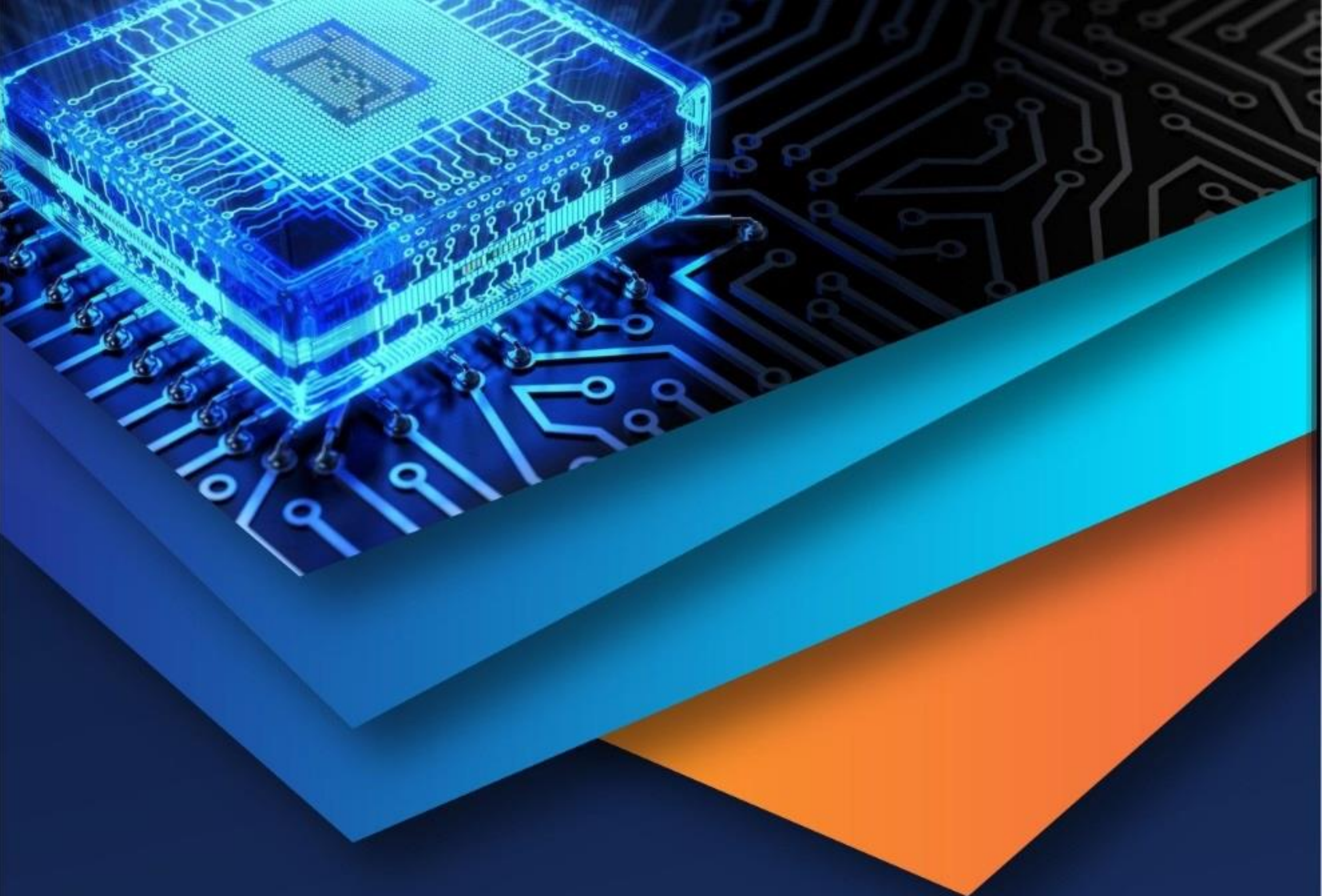
Visualization and Comp

[11] K. Karsch, C. Liu, and S. B. Kang. Depth extraction from video

using non-parametric sampling. In *Proc. European Conf. Computer*

Vision, 2012.

uter Graphics, 18(7):1079–1088, July 2012



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY