Diminution Reality Objects –based on Image Depth

Sallama Resen and Muthanna Ibrahim

Abstract--- Augmented Reality (AR) applications increasing popularity in a variety of industries, education, and marketing. AR techniques combined the real world with virtual objects enables Augmented Reality applications to provide a better understand and display of information for products. Diminished Reality Objects (DRO) techniques that visually eliminate real objects from AR applications. While an increasing interest in Diminution Reality technique can be observed that most of the Diminished reality research focuses on the consistency at the real–virtual and texture generated on a marker area. This paper handles the preservation depth consistency from edges and planar regions to build a depth map In order to develop DRO methods. the depth mask is built to this method has been consisting of two-stage run concurrence and each stage associated with error measuring to corrects stage. Results instantly which are Planarity and Boundary Depth techniques. Proposed method evaluated on RGB images dataset acquired by the digital camera with high-properties. The authenticity of the proposed method display in experimental results with a variety of criteria measurements.

Keywords--- Augmented Reality, Diminished Reality, Depth Map, Planar Regions.

I. INTRODUCTION

AR overlays real and virtual environment and displays them simultaneously on a mobile screen. This technology offers clients the ultimate imaginary and makes them feel the real experience. In AR technology virtual object placed in a real environment or real objects are extended with virtual objects while Diminished Reality is versa operation. Diminished Reality methods removing objects in a perceived environment in the way the observers assume that there is no apparent gap between the real and virtual scenes. With the increase of AR applications methods especially on the marketing business appear different techniques to solve AR applications faced it. Natural marker hiding built by considering two variable; texture and surrounding intensity of a marker area. These methods are inefficient for understanding depth scene in augmented reality applications. In DRO When hidden areas cannot be observed at all, high demand to address the problem of DRO. Traditional methods have no choice except to compensate for the background from the surrounding pixels without any depth feature consideration. Traditional methods classified by two categories of operations:

- Generating large image regions from sample textures used texture synthesis algorithms.
- Filling image gaps based on in painting techniques.

Image in painting referred to as image completion in a two-dimensional image plane. Hence, fairly reasonable results are probably positively rated, while the area around edges has clarified evident defects. Proposed methods aim to enhance traditional approaches and to be able to conclude depth information from images which are used predicted Depth maps for further 3D understanding scenarios [1]. Depth estimates area key requirement for augmented reality applications. Commonly used global statistics of depth used to estimate depth at the salient area,

Sallama Resen, Vocational Education, Ministry of Education. E-mail: sallamaresen@vocational.edu.iq

Muthanna Ibrahim, AL-Iraqi University, Collage of Art History Department. E-mail: muth1974a@yahoo.com

as plan surfaces. As this issue produced a novel DRO structure to overwhelmed deficits. Proposed DRO schema based depth maps built from combined features. From other side lack to datasets providing ground truth data for accurate evaluation. Dataset introduced a state-of-the-art method to evaluate using high-resolution RGB images dataset contain varied indoor scenarios to show performance measures.

In this part reviews literature concerning the features that could be applied in order to construct a depth map which make the DRO effective. Depth map reconstructed by projection mapping from 2D image points corresponding to the triangulation of 3D points projects from motion vectors [2]. Most prominently, methods based on the shape from shading principle [3, 4] exploit intensity or colour gradients from single images to predict the geometrical structure of objects under the assumptions of homogeneous lighting and reflection surface properties. While these methods operate passively on single images, active monocular alternatives have been investigated, such as, for instance, by exploiting the focus and defocus behaviour [5, 6] or polarization cues [9, 10]. With the emergence of light field cameras, a further line of approaches [7, 8] was developed. Approaches based on Data-driven learning proven to be effective than model-based methods [11]. One of the learning methods using for estimation depth maps was rough prediction points then refine the predictions [12] two networks used one for performing global points and other for local points[11].

A modified of this method uses combine predicts normal and semantic markers by deeper learning models [13] convolutional network combine conditional features used for preserving edges features which predicted depth gradients, Xu et al. [14] conditional random fields. In order to evaluate methods comprises marks originally designed for Multi-View Stereo obtained by a laser scanner [15] therefore, exhibit gaps in the depth map due to occlusions. Despite the possibility of getting a huge pair of images, mostly methods process a few images variation. Outdoor images suffer from a very low resolution associated depth maps built 3D scanner. However, are only provided in a very low indoor scene with aligned RGB and depth images obtained from video sequences serve for depth evaluation. DRO methods proposed to address with the numerous indoor scenarios. DRO propose depth map predicate have ability describing several plan surfaces. Established statistical depth metrics consider visual reviews for 3D point for depth maps. Edges and planar regions features are the perfect markers to calculate depth maps. The typical depth metrics are not able to assess the planarity of planar surfaces or the correctness of plane orientation estimation. These metrics have the ability to predicted depth maps for specifying on maps, different scene depths, and depth estimates consistently characteristics of the whole image which is not considered by the previous methods.

II. PROPOSED DEPTH MAP PREDICATION

The proposed method for removing objects from the AR scene based on the single image. Depth mask two-dimensional patterns focus on linear -dimensional patterns such as edge and object boundary. the depth mask is built to this method has been consisting by two-stage run concurrence and each stage associated with error measuring to corrects stage results instantly which are Planarity and Boundary Depth techniques. This paper presents a novel and efficient algorithm that combines the advantages of these two features to predicate the depth map as shown in figure(1).
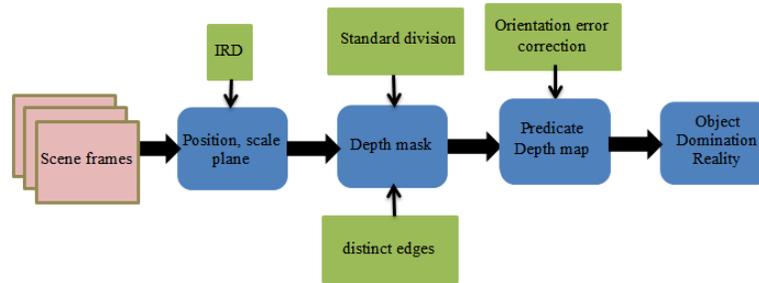
Figure 1: DRO Method Flowchart

The success of structure predicate depth map, however, is highly dependent on the error corrections which reflect the confidence in the created structure which is propagated to the next process. This procedure used to following described capture procedure gathered a dataset, which refers to them as the independent mark images dataset figure (2).
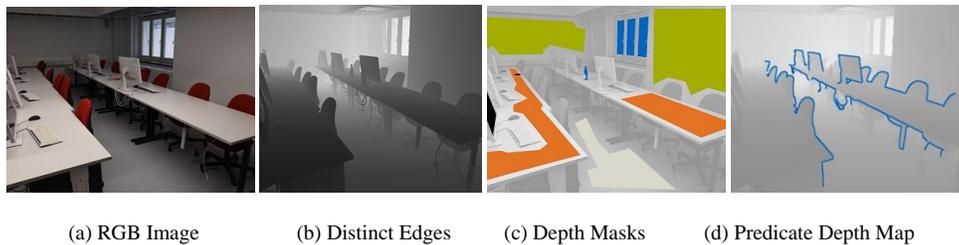
(a) RGB Image   (b) Distinct Edges   (c) Depth Masks   (d) Predicate Depth Map

Figure 2: The Proposed Method Samples Part

2.1 Convert Planarity

Global statistics are obtained over the included the image rate. Applying the definite range period by a quantity current depth series. Depth Range Interval (DRI) define slices permits into distinct bins one-meter depth to for close and far ranged objects to predicted depths. Objects assume planar constructions as floors, ceilings and walls. The shape correctness information cant determines from global statistics of objects in the scene, However, that permits considering the performance of predicted objects depths individualistically. Initialize set of marked images significant various surfaces construction to provide a set of rough structure {(Pmi , qmi)} superior position and scale are which Predicated by the following equation(1):

$$E(T) = \sum_i \left\| P_i^k - T_{qi}^k \right\|^2 \qquad (1)$$

The points are represented in homogeneous coordinates $T^m \in R4 \times 4$ is a comparison conversion equation (2):

$$T^k = \begin{bmatrix} cR & t \\ 0 & 1 \end{bmatrix} \qquad (2)$$

T^m represents object points in the image plane. Calculating planar object depths points still inspiring for many motives. Mainly, the objects vary by smooth colour gradients only which it is difficult to approximate the orientation correctness of a 3D plane. Correspondingly it is challenging to distinguish between a textured planar surface and a real depth discontinuity equation (3).

$$\pi_k = (\delta_k , d_k) \qquad (3)$$

depth correctness representation of planar constructions is a key mission for DRO adding the orientation for various planar

The flatness orientation to predicted 3D planes π_k depth map mask Y_k of a certain planar apparent is projected to 3D points $P_{k;\ i;j}$ where 3D planes π_k predicted 3D point clouds are fitted by equation(4).

$$\varepsilon_{PE}^{plan}(Y_K) = V\left[\sum_{P_{k,i,j}\in_{\overline{P}_k}} d\left(\pi_k, P_{k,i,j}\right)\right] \tag{4}$$

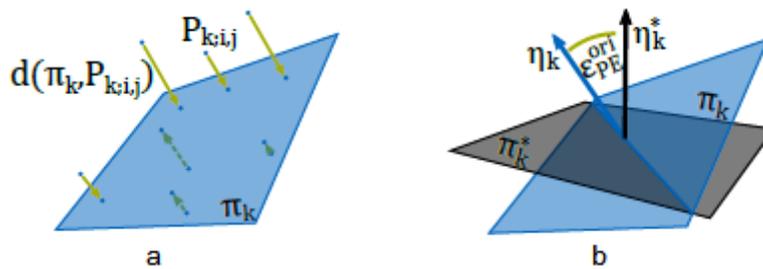the proposed planarity points and the planarity error Illustrate in figure (3).

Figure 3: Proposed Measurements for Planarity Surface

The orientation correction calculated from the average distance between the estimated 3D point and its matching 3-D plane evaluation equation(5).

$$\varepsilon_{PE}^{orie}(Y_K) = acos\left(\delta_k^T.\delta_k^*\right) \tag{5}$$

Figure(4) illustrate the planarity and depth boundary errors.

Note that the difference scaling planes used to build predicted depth maps, scaling can be removed according to the depth map points
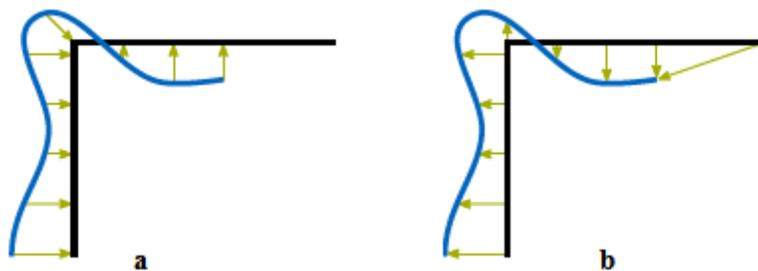

Figure 4: Planarity Errors (a) Depth Boundary Errors (b)

2.2 Depth Boundary

An indoor environment has a wide variety of depths object conditions in the sight. gaps objects depth is represented as gradient variations in the depth maps. In this research, evaluate generate depths maps that are characterized related depth points are continuities where avoid disorganized by texture to produce false depth discontinuities.

A depth discontinuities item preferred be stated by matching between edges in predicted depth and ground truth.

Two features are high standing for creating a ground truth depth transitions which are sharp edges and location accuracy in images. Sharp edges are taken from images using structured edges then selecting according to edges hypotheses order to correct.

Edges Y_{bin} are draw using structured edges then matching to the ground truth edges Y^*_{bin} with the binary edge by Euclidean distance equation (6).

$$E^* = DT(Y^*_{bin})$$

(6)

Threshold value given to ignored distances exceeding are. The depth boundary errors (DBEs) defined by accuracy equation (7)

$$\epsilon^{acc}_{DBE}(Y) = \frac{1}{\sum_i \sum_j y_{bin,i,j}} \sum_i \sum_j y_{bin,i,j} \cdot e^*_{i,j}$$

(7)

A completeness error defines to detect any absent edges in the predicted depth map equation (8).

$$\epsilon^{ccomp}_{DBE}(Y) = \frac{1}{\sum_i \sum_j y^*_{bin,i,j}} \sum_i \sum_j y^*_{bin,i,j} \cdot e_{i,j}$$

(8)

The DBEs is shown in Figure. 3a and Figure. 3b

III. DATASETS SETTINGS

Dataset images capture by a highly accurate camera, the high-quality camera allows capturing scene in high density with low noise. the experiment results in calibration usage a Nikon D7500 camera 200 mm and 35 mm focal length, 20.9 Megapixels, 4K UHD 3840 x 2160 at 30/25/24p. Selected manually a number of image points corresponding 2D, 3D to estimate the camera pose. The scene variety has features for statistics comparable, as the depth distribution thus, easy to test the robustness of DRO methods shown in figure (4). The last set comprises additional handheld images for every scene allowing validating multi-view images algorithms with good modality edges depth maps.

The dataset is collected for the evaluation depth maps produced by DRO methods. Samples of different scenes are provided in the dataset. A characterize of our dataset for evaluating DRO methods is provided in Table 1.

Table 1. Dataset Description

scenarios	Setting	Video length	Frame number	pixels
lecture room	simple	2.60sec	117	6016*4000
office room	simple	1.48sec	49	4000*6016
living rooms	simple	2.15sec	84	6016*4000
a factory room	simple	1.95sec	68	6016*4000
long corridors	complex	2.45sec	110	4000*6016
potted plants	complex	1.09sec	43	4000*6016
computer labs	simple	2.55sc	96	6016*4000

Several masks created manually samples masks are shown in figure (2), the main part of the dataset contains 7

scenes in total frames number 567. The dataset is accurate and wide-ranging indoor image dataset for depth predicate.

IV. EVALUATION OF EXPONENTS RESULTS

The proposed DRO methods evaluate the robustness of depth maps constructed by geometrical, colour transformations and the textured metrics for our reference dataset. Research with designs map on a planar surface exploit the most important features useful for DRO. Dataset images verified predicted correctly depth in the image without prior knowledge.

4.1 Standard Location Accuracy

The common error metrics are used to evaluate predicated depth:

$$\partial : \max\left(\frac{y_i}{y_i^*}, \frac{y_i^*}{y_i}\right) < th$$

$$absoulte\ real\ diffrence(rel) = \frac{1}{T}\sum_{i,j}|y_{,ji} - y_{i,j}^*| / y_{ij,}^*$$

$$squre\ real\ diffrence(sel) = \frac{1}{T}\sum_{i,j}|y_{i,j}^* - y_{i,j}|^2 / y_{ij,}^*$$

$$RMS\ (log) = \sqrt{\frac{1}{T}\sum_{I,J}|logy_{i,j} - logy_{i,j}^*|^2}$$

$$RMS\ (linear) = \sqrt{\frac{1}{T}\sum_{I,J}|y_{i,j} - y_{i,j}^*|^2}$$

Consequently, variations in the results are probable to appear from the established method report of the results of evaluating the DRO method displayed in the table. 2.

Table 2: Differences Metrics Results

scenarios	rel	see	RMS_{liner}	RMS_{log}
lecture room	0.36	0.12	2.94	0.22
office room	0.32	0.10	2.49	0.18
living rooms	0.26	0.06	2.84	0.17
a factory room	0.29	0.08	2.68	0.16
long corridors	0.27	0.07	2.54	0.18
potted plants	0.33	0.10	2.43	0.17
computer labs	0.25	0.06	2.96	0.14

In experiments, the proposed method applied Range Intervals fixed to 1 m RMS errors measure result displays a comparable tendency on datasets for the joint depth range. This proves the assumption of the lower mark creates from the enormous changes at 10 m depth range value. Proposed method have generalization capability which fulfils related results on images from a different camera with different intrinsic. Appearances error metrics used with the depth supplies shown in figure (5).
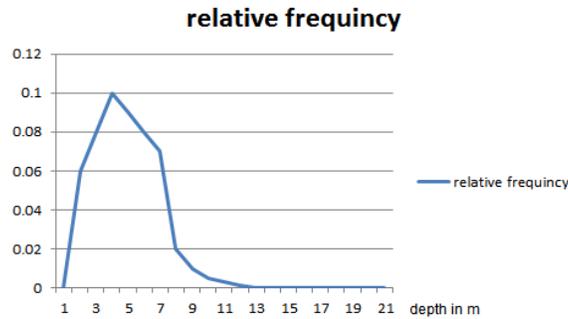
Figure 5: Distribution of Depth Values

4.2 Directed Depth Error

The Directed Depth Errors (DDEs) applied to provide information about the predicted depth is estimated too short or too far according to predicted and ground truth depths plan. The Directed Depth Errors (DDEs) define by equations:

$$\varepsilon_{DDE}^{+}(y) = \frac{|\{y_{i,j}|d_{sgn}(\pi,P_{i,j}) > 0 \wedge d_{sgn}(\pi,P^{*}_{i,j}) < 0\}|}{T}$$

$$\varepsilon_{DDE}^{-}(y) = \frac{|\{y_{i,j}|d_{sgn}(\pi,P_{i,j}) < 0 \wedge d_{sgn}(\pi,P^{*}_{i,j}) < 0\}|}{T}$$

$^{+}$DDE and $^{-}$DDE apply to detect the proportions distance of predicted depth pixels. Whether far or close a reference depth value π is fixed to 3m, figure(6) and table.4.

    (a) Original   (b) Estimates   (c) Depth Map   (d) Modifications
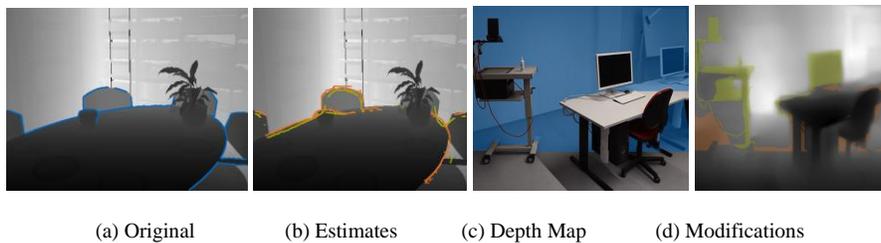
Figure 6: Visual Results after Applying DBE

Dataset reference enables a true valuation of depth gap through edges structure which computed the accuracy and completeness errors $^{"acc}$DBE and $^{"comp}$DBE, respectively, introduced in Section 2.2 results are recorded in Table 3. true depth boundaries produce sharp edges show in Figure 7 while missing edges stated by high values for $^{"comp}$DBE.

Table 3: Quantitative Results for the Proposed Method of Applying different Metrics

scenarios	ε_{PE}^{planc}	ε_{PE}^{orie}	ε_{DBE}^{acc}	ε_{DBE}^{comp}	ε_{DDE}^{-}	ε_{DDE}^{+}
lecture room	0.18	33.27	3.60	4808.	32.31	3.15
office room	0.21	26.64	3.01	32.00	21.51	3.84
living rooms	0.17	21.64	3.16	27.47	23.44	1.46
a factory room	0.22	32.02	4.58	38.41	20.89	1.99
long corridors	0.22	31.90	2.32	16.85	16.38	2.35
potted plants	0.20	26.67	2.36	21.02	16.44	2.57
computer labs	0.18	30.15	4.18	35.69	18.77	3.46

Evolution Location Depth Boundaries which define comprised of an accuracy measure $^{+}$ DBE, $^{-}$ DBE table.3.

The depth maps introduce the high-quality version, provided. Planarity error examines the reconstructed planar structures quality for different scenarios $"^{plan}$ the planarity errors $"^{orie}_{PE}$, and orientation errors, as explained in Section (2.1).

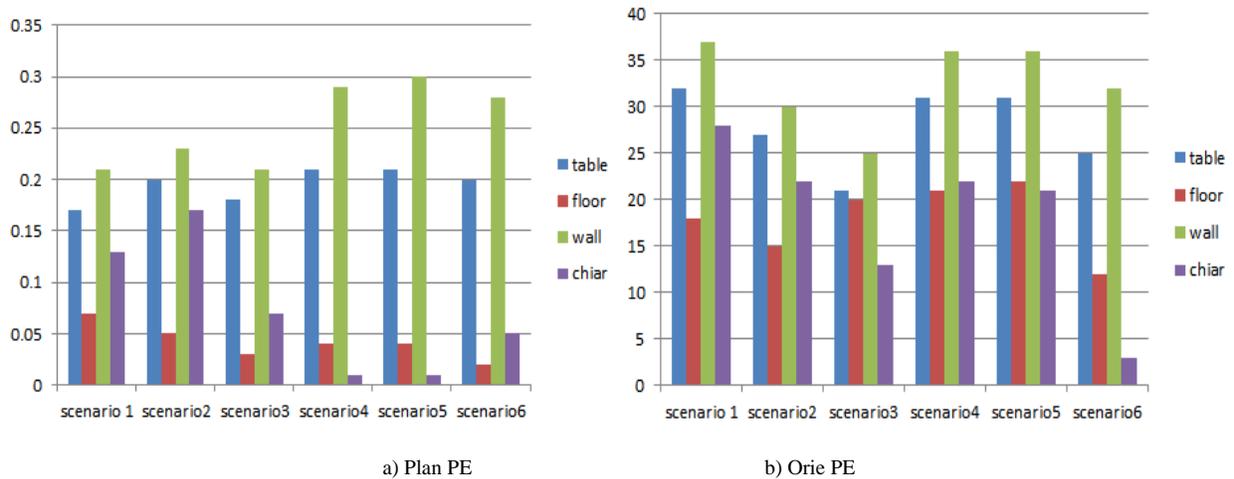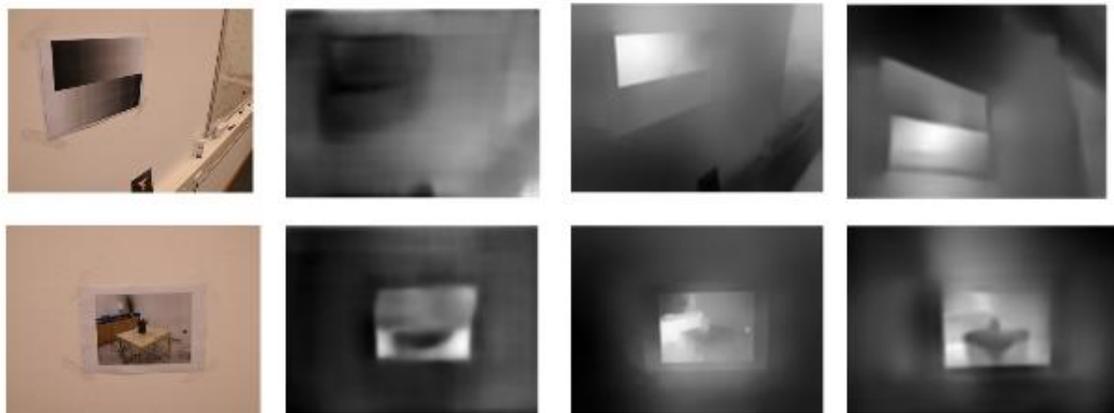a) Plan PE   b) Orie PE

Figure 7: Error Planarity Metrics

4.3 Data Augmentation

A set of augmented images extract from dataset to evaluate geometrical robustness of the proposed method. geometrical conversions overthrow the input images horizontally and vertically, which is probably to expose small appurtenances as shown in Figure.8.

(a) RGB (b) Color Transformations (c) Blurred Version (d) Gaussian Additive Noise

Figure 8: Predicted Depths for Two Samples from the Planar Surface

Dataset commonly illustrated pixels in the comedown part of the image, this is marked inspiration the estimated depth maps. global relative error metric applies to expressions the augmented images for DRO method results in Table 4.

Table 4: The Augmented Image Dataset Results

scenarios	input image	Camera Geometric Lens reflex		Contrast	hue	saturation
lecture room	0.360	-0024	0.068	0.059	0.010	-0.001
office room	0.318	-0.012	0.113	0.111	0.005	-0.001
living rooms	0.288	-0.017	0.137	0.110	0.002	-0.001
a factory room	0.274	-0.018	0.162	0.079	0.001	-0.001
long corridors	0.232	-0020	0.046	0.027	0.004	-0.001
potted plants	0.336	-0.014	0.131	0.031	0.011	-0.001
computer labs	0.248	-0.016	0.129	0..014	0.008	-0.001

V. CONCLUSIONS

Augmented Reality (AR) applications increasing popularity in a variety of industries, education, and marketing. Diminished Reality Objects (DRO) techniques that visually eliminate real objects from AR applications. This paper handles the preservation depth consistency from edges and planar regions to build a depth map In order to develop DRO methods. This paper presented a combination of statistics features to predicate depth map to develop DRO methods. Furthermore, capture a novel goodness dataset, favourable the absence ground truth information for proposed method experiments.

Experiments results show that evaluate the quality of the proposed approaches according to several metrics, like the edges, planar regions preservation, distance accuracy and depth uniformity. Commonly comparison used as error measurement, global statistical criteria proposed us to representation differences between DRO methods and the proposed method.

Reference Plane value established to recognize which depth values lie on the exact side and distinguishes between overestimated and underestimated predicted depths.

Experiment results obtained at 3 m a Reference Plane distance similarly the proportions computed for $^{0}_{DDE}$, overestimated $^{+}_{DDE}$, and underestimated $_{DDE}$ giving to the proposed error criteria .the properly and incorrectly predicted depths resulting, which are illustrated apparently that the statistical features be oriented to predict depths to short distance, estimated depths ranges to 90% and 80%. Quantitative results on the dataset exemplary listed for the global relative distance error. Errors showing relative differences for various image augmentations towards the error of the predicted original input image. In particular, our experiments have shown that the prediction of planar surfaces, which is crucial for many applications, is lacking accuracy.

REFERENCES

[1] Keisuke Tateno, Federico Tombari, Iro Laina, and Nassir Navab. Cnn-slam: Real-time dense monocular slam with learned depth prediction. arXiv preprint arXiv:1704.03489, 2017.
[2] Steven M. Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 519–528, 2006. doi:10.1109/CVPR.2006.19.
[3] Berthold K. P. Horn. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical report, MIT, Cambridge, MA, USA, 1970. URL https://dspace.mit.edu/handle/1721.1/6885.
[4] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 21(8):690–706, 1999.

[5]   Paolo Favaro and Stefano Soatto. A geometric approach to shape from defocus. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(3):406–417, 2005.doi: 10.1109/TPAMI.2005.43.

[6]   Supasorn Suwajanakorn, Carlos Hernandez, and Steven M.Seitz. Depth from focus with your mobile phone. *In IEEE Conference on Computer Vision and Pattern Recognition,* pages 3497–3506, 2015.

[7]   Andrea J. van Doorn, Jan J. Koenderink, and Johan Wagemans. Light fields and shape from shading. *Journal of Vision,* 11(3):21.1–21.21, 2011.

[8]   Stefan Heber and Thomas Pock. Convolutional networks for shape from light field. *In IEEE Conference on Computer Vision and Pattern Recognition,* pages 3746–3754, June 2016.

[9]   Trung Thanh Ngo, Hajime Nagahara, and Rin-ichiro Taniguchi. Shape and light directions from shading and polarization. *In IEEE Conference on Computer Vision and Pattern Recognition,* pages 2310–2318, 2015.

[10]   Achuta Kadambi, Vage Taamazyan, Boxin Shi, and Ramesh Raskar. Polarized 3d: High-quality depth sensing with polarization cues. *In IEEE International Conference on Computer Vision,* pages 3370–3378, 2015.

[11]   David Eigen and Rob Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. *In Proceedings of the IEEE International Conference on Computer Vision,* pages 2650–2658, 2015.

[12]   [12] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In International Conference on Neural Information Processing Systems, volume 2, pages 2366– 2374, 2014.

[13]   Bo Li, Yuchao Dai, Huahui Chen, and Mingyi He. Single image depth estimation by dilated deep residual convolutional neural network and soft-weight-sum inference. arXiv preprint arXiv:1705.00534, 2017.

[14]   Dan Xu, Elisa Ricci, Wanli Ouyang, Xiaogang Wang, and Nicu Sebe. Multi-scale continuous crfs as sequential deep networks for monocular depth estimation. arXiv preprint arXiv:1704.02157, 2017.

[15]   Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool,Pascal Fua, and Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. *In Computer Vision and Pattern Recognition,* 2008. CVPR 2008. *IEEE Conference on,* pages 1–8. Ieee, 2008.