Konstruksi: Publikasi Ilmu Teknik, Perencanaan Tata Ruang dan Teknik Sipil
Volume.3, Nomor.1 Tahun 2025
e-ISSN: 3031-4089; p-ISSN: 3031-5069, Hal 34-45
DOI: https://doi.org/10.61132/konstruksi.v3i1.685
Available online at: https://journal.aritekin.or.id/index.php/Konstruksi

Enhancing Depth Consistency in Augmented and Diminished Reality : Techniques and Evaluations Using RGB Imagery

Israa Shakir Seger[1*], Amjad Mahmood Hadi[2], Alaa Abd Ali Hadi[3]
[1]College of Basic Education, University of Muthanna, Muthanna, Iraq
[2]Al Muthanna University, Muthanna, Iraq
[3]Department Information and Communications Technology, Technical Institute of Samawah, Al-Furat Al-Awsat Technical University, Najaf, Iraq
Email : israa.shakir@mu.edu.iq, amjad@mu.edu.iq, alaa@atu.edu.iq

*Corresponding author: israa.shakir@mu.edu.iq**

Abstract : *Augmented Reality (AR) applications are rapidly gaining popularity across various industries, including education and marketing. By integrating real-world environments with virtual objects, AR enhances user understanding and information display for products. This paper explores Diminished Reality (DR) techniques, which are used to visually remove real objects from AR environments. Despite growing interest, much of the DR research predominantly focuses on maintaining consistency between real and virtual elements, particularly in texture handling on marker areas. Our study addresses the preservation of depth consistency using edge detection and planar segmentation to construct a depth map, essential for developing effective DR methods. We introduce a two-stage process involving depth mask construction, each stage equipped with error measurement for iterative refinement. Our proposed techniques, Planarity and Boundary Depth, are evaluated on a dataset of high-quality RGB images captured by digital cameras. Experimental results validate the effectiveness of our methods across various performance metrics, confirming the practicality of our approach in enhancing AR experiences.*

Keywords: *Augmented Reality, Diminished Reality, Depth Consistency, RGB Imaging, Error Measurement*

1. INTRODUCTION

Augmented Reality (AR) blends real and virtual environments on mobile screens, providing immersive experiences by integrating or augmenting real-world objects with virtual elements. Conversely, Diminished Reality (DR) techniques remove objects from the environment, seamlessly merging real and virtual scenes.[1] As AR expands in fields like marketing, new challenges arise, such as effectively hiding natural markers which consider both texture and light intensity. [2]Traditional DR approaches, which often overlook depth aspects, rely on either synthesizing textures for large areas or employing inpainting to fill gaps, but these methods struggle with edge defects. Our proposed improvements focus on incorporating depth information to refine these processes, utilizing predicted depth maps to enhance 3D understanding. Accurate depth estimation is vital, and our research introduces a comprehensive dataset of high-resolution RGB images for performance evaluation.[3], [4] We also review existing methods for constructing depth maps in DR, ranging from projection mapping and shape-from-shading techniques to advanced data-driven approaches using deep learning models.[5] These methods refine depth prediction and are assessed using a new metric that evaluates planarity and the

Received: Desember 07, 2024; Revised: Desember 21, 2024; Accepted: Januari 06, 2025; Published: Januari 08, 2025

accuracy of plane orientations, overcoming the limitations of previous techniques and providing more reliable results for both indoor and outdoor scenes.[6], [7]

2. SUGGESTED DEPTH MAP PREDICATION

The Suggested method aims to remove targets from AR scenes using a single image. This approach utilizes a depth mask that emphasizes two-dimensional linear patterns, such as edges and object boundaries.[8] The method operates through a two-stage concurrent process, where each stage incorporates error measurement to instantaneously correct and refine the outcomes using Planarity and Boundary Depth techniques.[9] Our research paper introduces an innovative and efficient algorithm that leverages these characteristics to predict the depth map as illustrated in Figure (I).

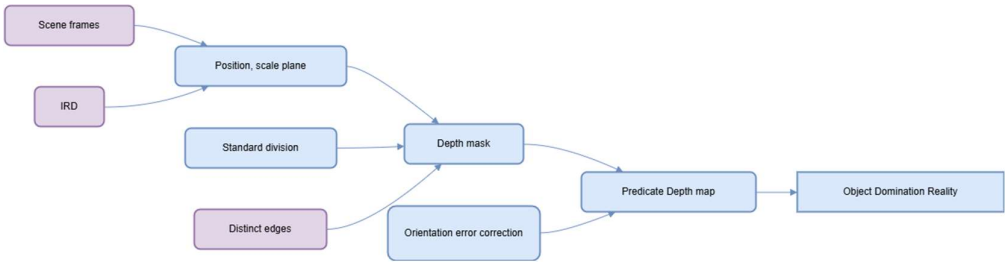


Figure 1. DRO method Diagram

The accuracy of the predicted depth map largely depends on the error correction mechanism, which enhances the reliability of the resultant structure.[10] This corrected structure then informs subsequent processes. The data for this study was collected using a specific capture procedure, resulting in a dataset referred to as the independent mark images dataset, shown in Figure II. This structured approach ensures that each phase of depth map prediction is robust and reliable, providing confidence in the generated depth maps.[11], [12]

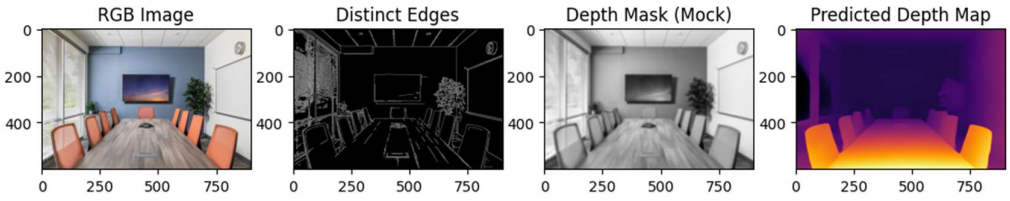


Figure 2. The Suggested method samples part

Convert Planarity

The proposed method involves gathering global statistics across the image at different depths by defining the Depth Range Interval (DRI). The DRI slices the depth data into distinct bins, each representing a one-meter depth interval to categorize objects based on their proximity—ranging from close to distant. These objects, typically planar structures such as floors, ceilings, and walls, cannot have their shape accuracy solely determined from global statistics[13]. However, such metrics allow for the individualistic performance evaluation of the predicted depths of these objects. Initially, a set of marked images representing various surface structures is utilized to establish a rough structure denoted by $\left(P_{\{mi\}}, q_{\{mi\}}\right)$ where the place and scale are predicted by the next equation (1):

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

Here, the points are represented in homogeneous coordinates. $(Tm \in R4 * 4)$ is a transformation matrix, represented by equation (2):

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

T^m maps the object points from 3D space to the image plane. Calculating the depth of planar objects is challenging, primarily because objects that vary by smooth color gradients complicate the accurate approximation of a 3D plane's orientation. Additionally, distinguishing between a textured planar surface and a real depth discontinuity is difficult, represented by equation (3):

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

The representation of depth accuracy and the orientation of planar constructions are crucial for effective DRO. The orientation and flatness of predicted 3D planes π_k are essential components. The depth map mask Y_K of a specific plane is projected onto 3D points $P_K; i; j$ where the fitted 3D point clouds are described by equation (4):

$$\varepsilon_{PE}^{plan}(Y_K) = V\left[\sum_{P_{k,i,j} \Leftarrow P_k} d\left(\pi_k, P_{k,i,j}\right)\right] \qquad (4)$$

The proposed planarity points and the associated errors are illustrated in Figure III.

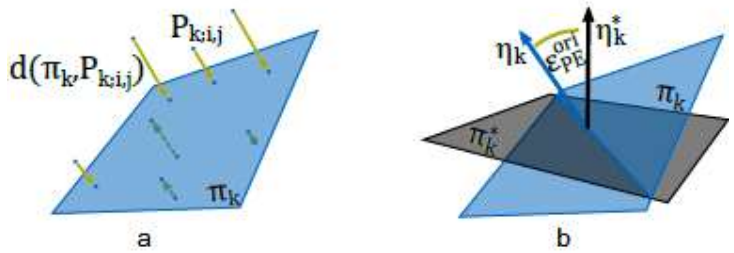


Figure 3. Suggested Measurements for Surface Planarity

This equation calculates the planarity error, which is illustrated in Figure 3. The orientation correction is derived from the average distance between the estimated 3D point and its corresponding 3D plane, as shown in equation (5):

$$\varepsilon_{PE}^{orie}(Y_K) = acos(\delta_k^T . \delta_k^*) \qquad (\mathbf{5})$$

Depth Boundary

Indoor environments feature complex depth conditions, represented as gradient changes on depth maps. This research evaluates depth maps that maintain continuous depth points, avoiding texture-induced false discontinuities.[14], [15]

Depth discontinuities are identified by comparing edges in predicted depth maps with ground truth, focusing on sharp edges and precise location accuracy derived from structured edge detection.[16] Edges in the $Y^*\ bin$ are matched to the ground truth $Y^*\ bin$ using binary edge comparison and Euclidean distance, as detailed in Equation (6).

$$E^* = \mathrm{DT}(Y_{bin}^*) \qquad (6)$$

Threshold values are set to ignore distances that exceed the specified limits. Depth boundary errors (DBEs) are defined according to the accuracy formula in Equation (7).

$$\epsilon_{DBE}^{acc}(Y) = \frac{1}{\sum_i \sum_j y_{bin,i,j}} \sum_i \sum_j y_{bin,i,j} \cdot e_{i,j}^* \qquad (7)$$

A completeness error is defined to identify any missing edges in the predicted depth map, as specified in Equation (8).

$$\epsilon_{DBE}^{ccomp}(Y) = \frac{1}{\sum_i \sum_j y_{bin,i,j}^*} \sum_i \sum_j y_{bin,i,j}^* \cdot e_{i,j} \qquad (8)$$

3. DATASETS SETTINGS

The dataset was captured using a high-precision Nikon D7500 camera with 200 mm and 35 mm lenses, 20.9 Megapixels, and 4K UHD video capability. Images were meticulously shot in high resolution with minimal noise, and specific points were manually selected to estimate the camera's 2D and 3D pose.[17] This setup facilitated testing the robustness of Diminished Reality (DRO) methods by providing diverse scene statistics like depth distribution, as shown in Figure (4).[18] Additionally, the dataset includes handheld images for each scene, supporting the validation of multi-view image algorithms that enhance depth map edges. This collection is specifically compiled to assess depth maps produced by DRO methods, with various scene samples and their characteristics outlined in Table I.

Table 1. dataset description

Script	Video length in Seconds	Frame No.	No. of pixel
Desk Room	2.60 sec	117	6016x4000
Living Room	1.48 sec	49	4000x6016
Lab Room	2.15 sec	84	6016x4000
The Lobby	1.95 sec	68	6016x4000
Plants	2.45 sec	110	4000x6016
PC Lab	1.09 sec	43	4000x6016
Session Room	2.55 sec	96	6016x4000

Several manually created masks are displayed in Figure II.[19] The primary component of the dataset includes a total of 7 scenes with 567 frames. This dataset is a comprehensive and accurate collection of indoor images for depth prediction.

4. ASSESSMENT OF EXPONENTS RESULTS

The suggested DRO methods assess the robustness of depth maps, which are constructed using geometric and color transformations, as well as textured metrics for our reference dataset. Our research focuses on designing maps on planar surfaces to exploit key features beneficial for DRO. The dataset images confirm the accurate prediction of depth in images without requiring prior knowledge.[20]

Benchmark Location accuracy

Benchmark error metrics are employed to assess the predicted depth :

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

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

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

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

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

As a result, variations in the outcomes are likely to emerge from the reported results of the DRO method evaluation, as shown in Table II.

Table 2. differences metrics results

Session Room	rel	see	RMSliner	RMSlog
Desk Room	0.35	0.13	2.95	0.21
Living Room	0.33	0.09	2.49	0.18
Lab Room	0.27	0.07	2.85	0.19
The Lobby	0.28	0.08	2.68	0.16
Plants	0.26	0.09	2.55	0.18
PC Lab	0.34	0.10	2.43	0.17
Session Room	0.26	0.06	2.96	0.15

In the experiments, the proposed method utilized fixed Range Intervals of 1m and measured RMS errors, which demonstrated consistent trends across datasets for the combined depth range. This supports the hypothesis that significant variations occur at a depth range of 10m. The proposed method exhibits generalization capabilities, producing consistent results on images taken from different cameras with varied intrinsic properties. Error metrics related to depth are presented in Figure V .[21], [22]

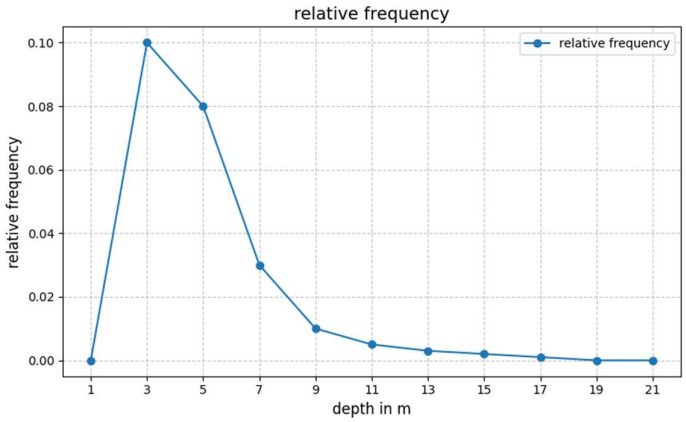


Figure 4. Allocation of depth values

Guided Depth Error

The Guided Depth Errors (DDEs) are used to indicate whether the predicted depth is estimated as too shallow or too deep compared to the actual depth as per the ground truth plan. The Guided Depth Errors (DDEs) are defined by the following equations:

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

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

The +DDE and -DDE metrics are utilized to measure the deviation of predicted depth pixels from a reference depth. Specifically, these metrics determine whether the depths are farther or closer than a set reference depth of 3m, as shown in Figure V and Table IV.[23]

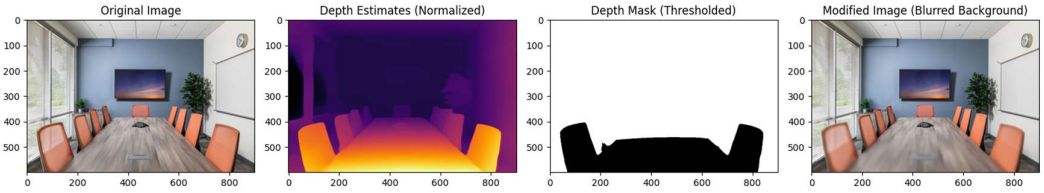


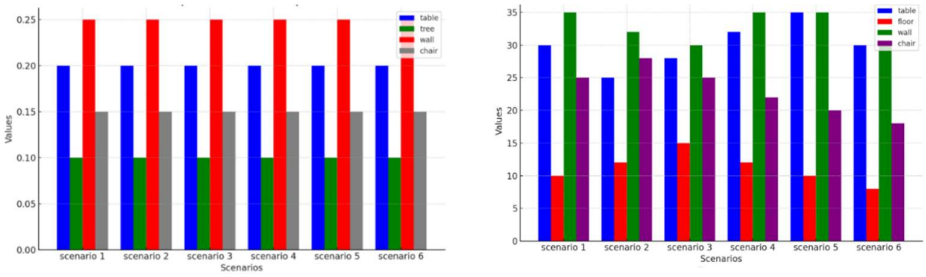
Figure 5. Visual outcomes following the application of DBE

The dataset provides a reliable means of assessing depth discrepancies along edge structures by calculating accuracy and completeness errors, denoted as acc_{DBE} and $comp_{DBE}$ respectively, as discussed in Section (2.2). The results are documented in the Table III.[24] True depth boundaries, which produce sharp edges, are displayed in Figure VII, while the presence of Absent edges is indicated by high values for $comp_{DBE}$.

Table 3. Quantitative outcomes from implementing various metrics with the proposed method

Session Room	ε_{PE}^{planc}	ε_{PE}^{orie}	ε_{DBE}^{acc}	ε_{DBE}^{comp}	ε_{DDE}^{-}	ε_{DDE}^{+}
Desk Room	0.18	33.27	3.60	48.08	32.31	3.15
Living Room	0.21	26.64	3.01	32.00	21.51	3.84
Lab Room	0.17	21.64	3.16	27.47	23.44	1.46
The Lobby	0.22	32.02	4.58	38.41	20.89	1.99
Plants	0.22	31.90	2.32	16.85	16.38	2.35
PC Lab	0.20	26.67	2.36	21.02	16.44	2.57
Session Room	0.18	30.15	4.18	35.69	18.77	3.46

The Evaluation of Location Depth Boundaries includes measures of accuracy such as +DBE and –DBE, detailed in Table 3. High-quality versions of the depth maps are provided.[25] The quality of reconstructed planar structures across various scenarios is assessed through Planarity Error, denoted as "^{plan} and Orientation Error, referred to as "$^{orie}_{PE}$ detailed in Section 2.1.



I - Plan [PE] II - ORIE [PE]

Figure 6. Measurement of non-planarity errors

Data Enhancement

A collection of enhanced images was extracted from the dataset to assess the geometric stability of the proposed approach. Geometrical transformations, including horizontal and vertical flipping of the input images, were applied, likely revealing minor details as depicted in Figure VII.

Figure 7. Predicted depths

The dataset frequently displays pixels in the lower portion of the image, which significantly influences the estimated depth maps.[26], [27] A global relative error metric is used to evaluate the enhanced images for the DRO method, with results presented in Table IV.

Table 4. the enhanced image dataset results

scenarios	input image	Camera Geometric Lens reflex		Contrast	hue	saturation
Session Room	0.360	-0024	0.059	0.010	-0.001	-0.001
Desk Room	0.318	-0.012	0.111	0.005	-0.001	-0.001
Living Room	0.288	-0.017	0.110	0.002	-0.001	-0.001
Lab Room	0.274	-0.018	0.079	0.001	-0.001	-0.001
The Lobby	0.232	-0020	0.027	0.004	-0.001	-0.001
Plants	0.336	-0.014	0.031	0.011	-0.001	-0.001
PC Lab	0.248	-0.016	0..014	0.008	-0.001	-0.001

5. CONCLUSIONS

Augmented Reality (AR) applications are expanding across industries like education and marketing, using Diminished Reality Objects (DRO) to visually remove real objects from AR settings. This paper emphasizes maintaining depth consistency in edges and planar areas to develop effective DRO methods, introducing statistical features for depth map prediction alongside a novel dataset that compensates for the absence of ground truth data. Experimental results evaluate the proposed methods against metrics such as edge preservation, accuracy of distance, and depth uniformity. A Reference Plane value helps differentiate accurately estimated depths from those that are over or underestimated. The results show that while the prediction of short distances achieves high accuracy (90% and 80%), the accuracy in predicting planar surfaces crucial for many applications remains inadequate. The experiments highlight relative errors across different image augmentations, underscoring the need for improvements in planar surface predictions.

References

Cheng, Y. F., Yin, H., Yan, Y., Gugenheimer, J., & Lindlbauer, D. (2022). Towards understanding diminished reality. In *Conference on Human Factors in Computing Systems - Proceedings* (pp. 3517452). Association for Computing Machinery. https://doi.org/10.1145/3491102.3517452

Chua, T. S., Ngo, C.-W., Kumar, R., Lauw, H. W., & Lee, R. K.-W. (2024). Companion proceedings of the ACM Web Conference 2024 (WWW '24 Companion): May 13-17, 2024, Singapore, Singapore. The Association for Computing Machinery.

Dhamo, H., Navab, N., & Tombari, F. (n.d.). Object-driven multi-layer scene decomposition from a single image.

Dhamo, H., Tateno, K., Laina, I., Navab, N., & Tombari, F. (2018). Peeking behind objects: Layered depth prediction from a single image. *Pattern Recognition Letters, 132*, 191–202. https://doi.org/10.1016/j.patrec.2019.05.007

Eskandari, R., & Motamedi, A. (n.d.). Diminished reality in architectural and environmental design: Literature review of techniques, applications, and challenges.

Fujii, R., Hachiuma, R., & Saito, H. (2021). RGB-D image inpainting using generative adversarial network with a late fusion approach. Retrieved from http://arxiv.org/abs/2110.07413

Gkitsas, V., Sterzentsenko, V., Zioulis, N., Albanis, G., & Zarpalas, D. (n.d.). PanoDR: Spherical panorama diminished reality for indoor scenes.

Hadi, A. A. A., & Hadi, A. M. (2024). Improving cybersecurity with random forest algorithm-based big data intrusion detection system: A performance analysis. In *AIP Conference Proceedings* (Vol. 232). American Institute of Physics. https://doi.org/10.1063/5.0191707

Hadi, A. M. (2024). Enhancing MRI brain tumor classification with a novel hybrid PCA+RST feature selection approach: Methodology and comparative analysis. *International Journal of Computational and Electronic Aspects in Engineering, 5*(2), 116–130. https://doi.org/10.26706/ijceae.5.3.20240805

Kán, P., & Kafumann, H. (2019). DeepLight: Light source estimation for augmented reality using deep learning. *Visual Computer, 35*(6–8), 873–883. https://doi.org/10.1007/s00371-019-01666-x

Kari, M., et al. (n.d.). TransforMR: Pose-aware object substitution for composing alternate mixed realities. Retrieved from https://developer.apple.com/documentation/arkit

Kikuchi, T., Fukuda, T., & Yabuki, N. (n.d.). Automatic diminished reality-based virtual demolition method using semantic segmentation and generative adversarial network for landscape assessment.

Kobayashi, K., & Takahashi, M. (2024). Real-time diminished reality application specifying target based on 3D region. *Virtual Worlds, 3*(1), 115–134. https://doi.org/10.3390/virtualworlds3010006

Kulshreshtha, P., Lianos, K.-N., Pugh, B., & Jiddi, S. (2022). Layout-aware inpainting for automated furniture removal in indoor scenes. Retrieved from http://arxiv.org/abs/2210.15796

Li, P., Liu, L., Schönlieb, C.-B., & Aviles-Rivero, A. I. (2024). Optimised ProPainter for video diminished reality inpainting. Retrieved from http://arxiv.org/abs/2406.02287

Li, Z., et al. (2022). Color-to-depth mappings as depth cues in virtual reality. In *UIST 2022 - Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (pp. 3545646). Association for Computing Machinery. https://doi.org/10.1145/3526113.3545646

Ma, F., & Karaman, S. (2017). Sparse-to-dense: Depth prediction from sparse depth samples and a single image. Retrieved from http://arxiv.org/abs/1709.07492

Macedo, A. C. F., Apolinário, A. L., & Jr, A. (2015). Occlusion handling in augmented reality: Past, present, and future.

Mansoor, R., Sasse, H., Ison, S., & Duffy, A. (2015). Crosstalk bandwidth of grating-assisted ring resonator add/drop filter. *Optical and Quantum Electronics, 47*(5), 1127–1137. https://doi.org/10.1007/s11082-014-9969-0

Ming, Y., Meng, X., Fan, C., & Yu, H. (n.d.). Deep learning for monocular depth estimation: A review.

Mori, S., Ikeda, S., & Saito, H. (2017). A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Transactions on Computer Vision and Applications, 9*(1). https://doi.org/10.1186/s41074-017-0028-1

Mori, S., Shibata, F., Kimura, A., & Tamura, H. (2015). Efficient use of textured 3D model for pre-observation-based diminished reality. In *Proceedings of the 2015 IEEE International Symposium on Mixed and Augmented Reality Workshops, ISMARW 2015* (pp. 32–39). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ISMARW.2015.16

Ramamonjisoa, M., & Lepetit, V. (n.d.). SharpNet: Fast and accurate recovery of occluding contours in monocular depth estimation.

Valentin, J., et al. (2018). Depth from motion for smartphone AR. In *SIGGRAPH Asia 2018 Technical Papers* (pp. Article 3275041). Association for Computing Machinery. https://doi.org/10.1145/3272127.3275041

Wasenmuller, O., Meyer, M., & Stricker, D. (2016). Augmented reality 3D discrepancy check in industrial applications. In *Proceedings of the 2016 IEEE International Symposium on Mixed and Augmented Reality, ISMAR 2016* (pp. 125–134). Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ISMAR.2016.15

Yang, J., Ye, X., Li, K., Hou, C., & Wang, Y. (2014). Color-guided depth recovery from RGB-D data using an adaptive autoregressive model. *IEEE Transactions on Image Processing, 23*(8), 3443–3458. https://doi.org/10.1109/TIP.2014.2329776

Zhang, Y., Scargill, T., Vaishnav, A., Premsankar, G., Di Francesco, M., & Gorlatova, M. (2022). InDepth: Real-time depth inpainting for mobile augmented reality. *Proceedings of the ACM Interactive Mobile, Wearable, and Ubiquitous Technologies, 6*(1), Article 10. https://doi.org/10.1145/3517260