DOI: https://doi.org/10.15276/hait.07.2024.25
UDC 004.93

Robust depth map refining using color image

Sergey B. Kondratyev[1)]
ORCID: https://orcid.org/0000-0003-4975-5757; voshodvostok@gmail.com
Svitlana G. Antoshchuk[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@op.edu.ua. Scopus Author ID: 8393582500
Mykola A. Hodovychenko[1)]
ORCID: https://orcid.org/0000-0001-5422-3048; hodovychenko@od.edu.ua. ScopusAuthor ID:57188700773
Serhii A. Ustenko[1)]
ORCID: https://orcid.org/0000-0003-4968-1233; ustenko.s.a@gmail.com. Scopus Author ID:57207577774
[1)] Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ABSTRACT

Depth maps are essential for various applications, providing spatial information about object arrangement in a scene. They play a crucial role in fields such as computer vision, robotics, augmented and virtual reality, autonomous systems, and medical imaging. However, generating accurate, high-quality depth maps is challenging due to issues like texture-copying artifacts, edge leakage, and depth edge distortion. This study introduces a novel method for refining depth maps by integrating information from color images, combining structural and statistical techniques for superior results. The proposed approach employs a structural method to calculate affinities within a regularization framework, utilizing minimum spanning trees (MST) and minimum spanning forests (MSF). Super-pixel segmentation is used to prevent MST construction across depth edges, addressing edge-leaking artifacts while preserving details. An edge inconsistency measurement model further reduces texture-copying artifacts. Additionally, an adaptive regularization window dynamically adjusts its bandwidth based on local depth variations, enabling effective handling of noise and maintaining sharp depth edges. Experimental evaluations across multiple datasets show the method's robustness and accuracy. It consistently achieves the lowest mean absolute deviation (MAD) compared to existing techniques across various upsampling factors, including 2×, 4×, 8×, and 16×. Visual assessments confirm its ability to produce depth maps free of texture-copying artifacts and blurred edges, yielding results closest to ground truth. Computational efficiency is ensured through a divide-and-conquer algorithm for spanning tree computations, reducing complexity while maintaining precision. This research underscores the importance of combining structural and statistical information in depth map refinement. By overcoming the limitations of existing methods, the proposed approach provides a practical solution for improving depth maps in applications requiring high precision and efficiency, such as robotics, virtual reality, and autonomous systems. Future work will focus on real-time applications and integration with advanced depth-sensing technologies.

Keywords: Depth maps; 3D reconstruction; image processing; spatial data analysis; data refinement; sensor-based imaging; edge detection; noise reduction; depth sensing; computational imaging; augmented reality; autonomous systems

***For citation*: Kondratyev S. B., Antoshchuk S. G., Hodovychenko M. A. Ustenko S. A. "Robust depth map refining using color image". *Herald of Advanced Information Technology.* 2024; Vol.7 No.4: 361–370. DOI: https://doi.org/10.15276/hait.07.2024.25**

INTRODUCTION, FORMULATION OF THE PROBLEM

Depth maps play a crucial role in various fields, as they provide information about the spatial arrangement of objects in a scene. They enable machines and systems to understand the three-dimensional structure of the environment, which is particularly important for tasks in computer vision and robotics [1].

Depth maps make it possible to create more realistic and interactive applications, such as those in augmented and virtual reality.

These data are also widely used in machine learning for tasks like object detection, scene segmentation, and pose estimation, enhancing the accuracy and reliability of models.

In fields such as medicine, autonomous vehicles, gaming, and cartography, depth maps are applied to data analysis, navigation, and visualization of complex objects and environments.

Depth maps can be generated using a variety of methods depending on the application and available resources. Active sensing methods involve emitting signals into the environment and analyzing their responses. For example, stereo vision uses two cameras to capture different perspectives of a scene and calculates depth by comparing disparities between the images. LiDAR emits laser pulses and measures the time it takes for the reflected light to return, while time-of-flight cameras determine depth by measuring the travel time of light to and from an object. Structured light techniques project known patterns onto a scene and analyze their distortion to estimate depth [2]. On the other hand, passive sensing methods rely on analyzing existing visual data without emitting signals.

Monocular depth estimation uses a single image and algorithms, often powered by machine learning, to infer depth from visual cues such as shading and texture gradients.

Multi-view stereo processes multiple images taken from different viewpoints to reconstruct a scene's depth. Both active and passive methods are widely used in various fields, such as robotics, virtual reality, and autonomous systems, to provide accurate spatial information [3].

Active methods for generating depth maps face several challenges that can affect their performance and applicability. These shortcomings can be addressed to some extent using enhancement techniques that utilize color images, as they offer supplementary structural details about the scene. The close relationship between texture and depth variation enables notable enhancements in the precision of depth maps. As a result, combining depth information with color data emerges as a compelling approach for improving depth sensing systems [4].

Thus, **the purpose of this study** is to develop a method for robust depth map refining with the help of color image.

The proposed method better retains the structure of the depth map by using a structural approach to calculate affinities in the regularization term.

The proposed method differs from other methods in that it allows for more accurate calculation of affinities by utilizing the space formed by multiple spanning trees.

The proposed affinity calculations are grounded on paths within a spanning tree or several neighboring spanning trees. This approach will allow an efficient representation of the local depth structure. In addition, the edge weights of each spanning tree represent a soft boundary dissimilarity metric, which reduces texture duplication artifacts.

1. LITERATURE REVIEW

The existing methods, including those proposed in the previous subsections, compute guidance affinities in the regularization term using a non-structural approach based solely on color and depth differences between a pixel and its neighbors [5, 6], [7].

This approach neglects the local structure of the depth map, which can result in excessive smoothing of depth edges on the enhanced depth map, particularly in cases like 16× depth super-resolution [8].

In contrast, it has been observed that the tree filter demonstrates more robust performance in preserving edges. This filter is widely used in various computer vision tasks, such as structure-preserving smoothing and stereo matching. In these methods, an MST (Minimum Spanning Tree) is utilized to automatically separate dissimilar pixels that are spatially close to each other, turning the tree distance into an edge-aware metric [9].

However, global MSTs suffer from the issue of edge leakage, where a single tree connects all pixels in an image, spreading smoothing effects globally along the MST. This construction inevitably crosses strong edges, leading to their distortion [10].

The proposed method calculates anisotropic guidance affinities in the regularization term based on the tree distance between two pixels. To avoid constructing an MST across depth edges, which causes edge-leaking artifacts, the method introduces an approach inspired by creating an MST for each super-pixel on a color image generated through over-segmentation [11].

Since the color image is high-quality, and pixels within the same super-pixel share similar characteristics, tree distance calculations become more reliable within super-pixel regions. Moreover, due to over-segmentation, pixels and their neighbors located on adjacent MSTs may have similar depth values, so connections between adjacent MSTs must also be considered. These neighboring MSTs are combined into a structure referred to as the Minimum Spanning Forest (MSF) [12]

If guidance affinities are computed based only on the registered color image, this can lead to texture-copying artifacts and blurred depth edges caused by inconsistencies between the depth map and the registered color image [13].

Building on the proposed inconsistency model, which has proven effective in mitigating texture-copying artifacts, edge weights within each MST in the MSF are calculated based on this model. These two complementary components work together to achieve robust depth map enhancement.

2. PROPOSED METHOD

2.1. Generation of minimum spanning tree

The super-pixel segmentation algorithm is utilized to perform over-segmentation on the color

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

image, effectively preventing the construction of a minimum spanning tree across depth edges and thereby reducing leaking-edge artifacts [14].

This advanced segmentation technique offers strong alignment with prominent image edges while maintaining linear computational complexity. Within each super-pixel, an 8-connected weighted subgraph $SG(N, E, W)$ is created.

The nodes N of this subgraph represent all the pixels within the segment, E includes all edges connecting these nodes, and W denotes the set of edge weights, defined as per the specified equation:

$$W(k, l) = \left|\nabla_c^{kl}\right|, \quad (1)$$

where ∇_c^{kl} is the color variation of k and l. Since the color image and the depth map exhibit different texture patterns, this configuration results in texture-copying artifacts and blurred depth edges.

To address these issues, the proposed method explicitly incorporates an edge inconsistency measurement into the construction of the spanning tree. Currently, within each spanning tree, the weight of an edge connecting two neighboring nodes is determined by their color similarity, as defined in equation (1).

Using these subgraphs, a spanning tree is constructed for each super-pixel by eliminating redundant edges. However, due to over-segmentation, pixels and their neighboring pixels located on adjacent spanning trees may still share similar depth values. Consequently, it is necessary to account for edges that connect adjacent spanning trees [15].

To ensure consistency with the weights of edges within each tree, the weights of edges linking adjacent trees are calculated based on color and depth differences, thereby reducing texture-copying artifacts.

Specifically, edges defined by pixels with similar color and depth values, situated on adjacent spanning trees, extend the spanning tree of a super-pixel into a forest for the entire image without crossing significant edges [16].

This approach can be viewed as a supervisory mechanism to mitigate the leaking-edge problem. A similar configuration is used in methods that rely solely on color similarity for edge-aware smoothing. The proposed forest, however, can be interpreted as a distance space constructed based on image content, where spatial factors are integrated into the paths between pixels.

Unlike the non-structural Euclidean distance space employed in the dual kernels of bilateral filters, the distances computed in the proposed MSF reflect the structural characteristics of the image [17].

The proposed method represents each pixel p as a two-dimensional point $Pnt_p\left(clr_p, dv_p\right)$, where clr_p denotes its color value and dv_p represents its depth value. The distance metric between two pixels is defined using the L1-norm.

The optimal pair of nodes (k', l') connecting adjacent super-pixels SP_α and SP_β is determined as:

$$(k', l') = \min_{\substack{k \in SP_\alpha \\ l \in SP_\beta}} \|Pnt_k - Pnt_l\|, \quad (2)$$

when performing an exhaustive search, the computational complexity is $O(t_1, t_2)$, where t_1 and t_2 represent the number of nodes in their respective trees.

This approach becomes computationally expensive when the super-pixels are large. To overcome this issue, the proposed method employs a more efficient divide-and-conquer algorithm which reduces the complexity to $O\big((t_1 + t_2)\log(t_1 + t_2)\big)$.

The distance between two neighboring nodes in the regularization term is calculated as follows.

If k and l belong to the same super-pixel, their distance is determined using the standard tree distance along the path within the spanning tree:

$$d(k, l) = \sum_{i=0}^{n} W(k_i, k_{i+1}), \quad (3)$$

where the distance between adjacent nodes (k_i, k_{i+1}) along the path is computed efficiently using the Lowest Common Ancestor method.

If k and l are located in different super-pixels, the distance is computed in such components:

$$d(k, l) = d(p, p') + d(l, l') + 0.5 \times \left(\nabla_c^{k'l'} + \nabla_d^{k'l'}\right). \quad (4)$$

Here, k' and l' are the closest pair of nodes within adjacent trees, considering both color and depth similarities, as expressed in the final term of equation (4).

The terms $\nabla_c^{k'l'}$ and $\nabla_d^{k'l'}$ represent the color and depth differences between pixels k' and l' from the guided color image and the coarsely interpolated depth map, respectively. The pairs (p, p') and (l, l') correspond to pixels located within the same tree.

2.2. Using edge dissimilarity in spanning tree

The edge dissimilarity measurement model demonstrates a strong capability in reducing texture-copying artifacts.

To ensure clarity, this section summarizes the model as follows: it employs a bi-directional evaluation by alternating the roles of the color edge map and the depth edge map. In each direction, the optimal matching pairs of edge pixels are identified using forest optimization [18].

The data term incorporates local structural information, which is determined through Minimum Weighted Bipartite Matching, while the regularization term reflects global structural information.

The cost associated with the optimal matching of edge pixels serves as the edge inconsistency measurement, ranging from [0, 1].

Inspired by this model, the proposed method integrates it into the spanning tree construction to reduce texture-copying artifacts. In the method, θ represents a set of confidence values for pixels [19].

A smaller θ value indicates greater edge consistency. More specifically, the method redefines W for the edge (k, l) as follows:

$$W'(k,l) = \left|\nabla_c^{kl}\right| \times (1 - \theta_{kl}) + \left|\nabla_d^{kl}\right| \times \theta_{kl}, \quad (5)$$

where θ_{kl} is the confidence value for the pixel pair (k, l), is defined as $\theta_{kl} = \max\big(\theta(k), \theta(l)\big)$ to enhance the preservation of depth edges. Here $\theta(k)$ and $\theta(l)$ represent the confidence values for pixels k and l, respectively.

When the color edge map aligns more closely with the depth edge map (i.e., θ_{kl} approaches zero), ∇_c^{kl} takes on a more significant role in determining edge weights within the spanning tree, and vice versa.

The newly defined spanning edge weights W' replace the original weights [20].

Unlike earlier forest-based depth enhancement approaches, which do not incorporate structural information in regularization term calculations, the proposed method leverages the forest distance to better preserve depth edges.

Furthermore, texture-copying artifacts are substantially reduced by utilizing the improved spanning tree construction, which explicitly incorporates the edge inconsistency measurement model [21].

2.3. Window variations for calculating affinities

In [22], it was observed that the window ω has a significant impact on the model's performance. For example, when two pixels used in the regularization term of the forest exhibit markedly different depth values, a smaller ω provides superior results.

Consequently, if the depth difference between neighboring pixel pairs in the regularization term is known, the bandwidth ω can be adaptively selected as a prior to reconstruct a high-quality depth map. This adaptive approach helps preserve depth edges and reduces artifacts caused by noise and texture copying [23].

In this method, the coarsely interpolated depth map $\widehat{DM}$ is employed to estimate the prior, which dynamically controls ω.

However, due to the low quality of $\widehat{DM}$, the estimated edge locations may deviate from their true positions.

This makes it unreliable to directly calculate the depth difference between a pixel pair k and l. To address this, the proposed method defines two sub-regions (9×9 in size) centered at k and l, forming pixel sets Set_k and Set_l.

The maximum absolute difference $Diff_{max}$ between these sets is then used to estimate the potential prior for k and l.

A large $Diff_{max}$ indicates that the pixel pair is near a depth edge, and a small ω should be assigned to better preserve the edge. Conversely, a small $Diff_{max}$ suggests that the pixel pair is in a smooth region, where a larger ω can be chosen to suppress noise and minimize texture-copying artifacts [24].

The straightforward calculation of $Diff_{max}$ involves determining the differences for all elements in Set_k and Set_l, which has a computational complexity of $O(v^2)$, where v is the number of elements in each set. To improve efficiency, a method with $O(v)$ complexity is introduced.

First, the minimum and maximum depth values $(\max_k, \min_k, \max_l, \min_l)$ are computed for Set_k and Set_l. $Diff_{max}$ is then calculated as:

$$Diff_{max} = \max(|max_k - min_l|, |max_l - min_k|). \quad (6)$$

Given the variability in noise conditions, it is challenging to model the relationship between ω and $Diff_{max}$ with a fixed function (e.g., linear).

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

The proposed method calculates $Diff_{max}$ for each pixel pair in the regularization term and classifies them into three categories using double thresholds t_l and t_h, representing regions near strong edges, weak edges, and smooth areas. Different ω values are then assigned for each region as defined below:

$$\omega = \begin{cases} 2, & Diff_{max} > t_h (strong\ edge) \\ 10, & Diff_{max} < t_l (smooth\ area). \\ 4, & otherwise (weak\ edge) \end{cases} \quad (7)$$

3. EXPERIMENTAL RESULTS

The evaluation of the proposed method focuses on depth map super-resolution and depth map completion across various datasets. In the experiment conducted on synthetic datasets downsampling degradation was considered.

The noise-free datasets, based on ground truth data from the Middlebury Datasets, were downscaled using nearest neighbor interpolation [25].

The proposed method was tested at various upsampling factors, including 2×, 4×, 8×, and 16×, and compared against other methods. Tables 1 and 2 present the upsampling results for four different factors, with optimal results highlighted in bold.

The proposed method consistently achieves the lowest Mean Absolute Deviation in all cases.

The Mean Absolute Difference is a statistical indicator that computes the average of the absolute deviations between corresponding elements in two datasets. It is widely applied to measure the overall error or discrepancy between predicted and actual values or between two images in the field of image processing.

$$MAD = \frac{1}{N}\sum_{i=1}^{N}|x_i - y_i|, \quad (8)$$

where N represents the total number of elements in the datasets, and x_i and y_i are the corresponding elements from the two datasets being compared.

Thanks to the forest-based approach for computing guidance affinities, the proposed method attains the lowest MAD across all datasets.

Fig. 1 illustrates the experimental results for 8× upsampled depth maps on the "Laundry" and "Dolls" datasets.

The results indicate that Total Generalized Variation suffers significantly from texture-copying artifacts. Additionally, artifacts near depth edges are noticeable in the results of Joint Geodesic Filtering. In contrast, the proposed method produces results closest to the ground truth, with no texture-copying artifacts or blurred depth edges.

Table 1. **Experimental results of the proposed method and peer methods**

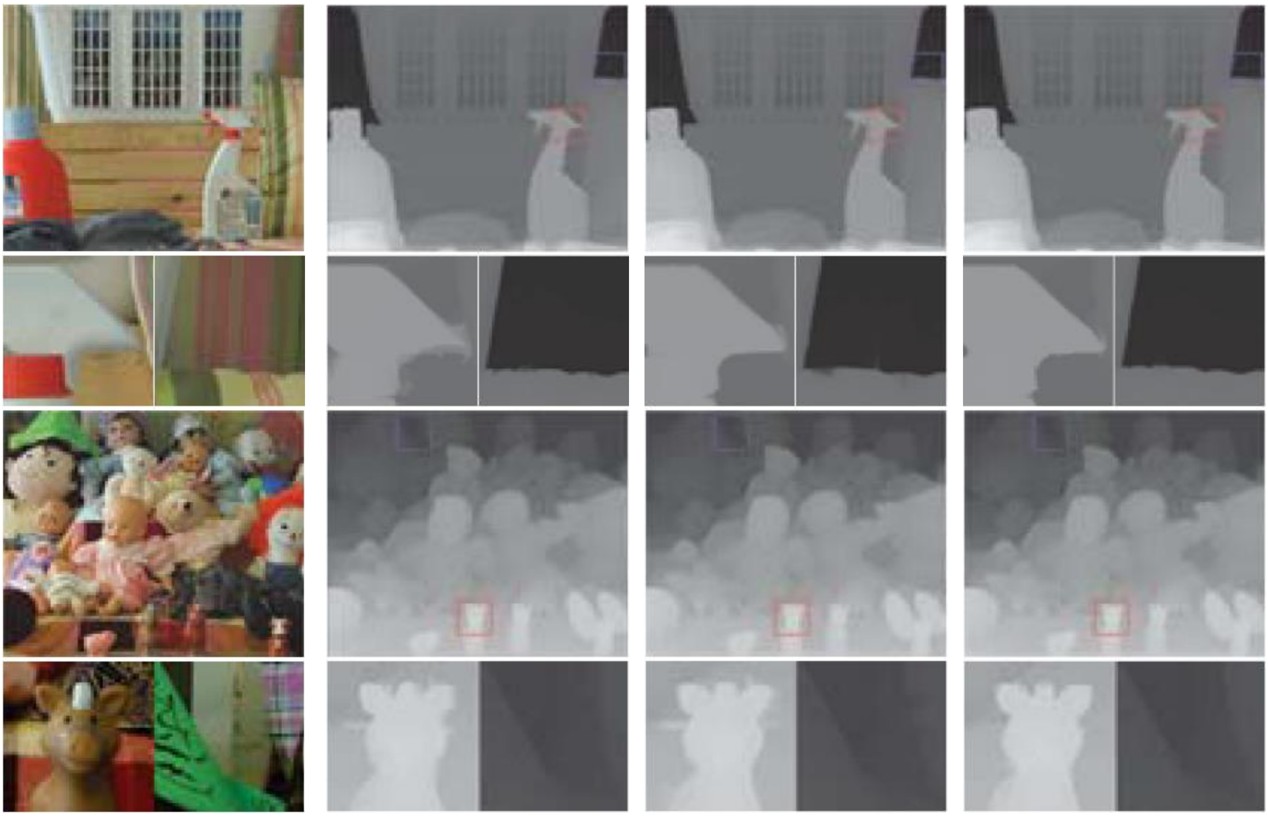
Method	Dataset											
	Book				Moebius				Art			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
AR [26]	0.13	0.21	0.36	0.78	0.13	0.23	0.42	0.83	0.18	0.49	0.64	1.99
Guided [27]	0.23	0.36	0.59	1.15	0.24	0.39	0.61	1.17	0.65	1.03	1.69	3.48
JBU [28]	0.18	0.37	0.75	1.58	0.19	0.38	0.77	1.48	0.46	0.86	1.70	3.37
TGV [29]	0.20	0.28	0.43	0.84	0.21	0.30	0.51	0.89	0.47	0.67	1.19	2.39
Bicubic [30]	0.14	0.30	0.61	1.16	0.14	0.32	0.61	1.15	0.50	1.00	1.88	3.61
MLS [31]	0.17	0.28	0.47	1.18	0.16	0.26	0.51	0.94	0.28	0.70	1.06	2.19
OMRF [32]	0.22	0.34	0.59	1.19	0.25	0.37	0.66	1.27	0.60	0.98	1.91	3.80
Proposed	**0.09**	**0.18**	**0.33**	**0.69**	**0.12**	**0.20**	**0.38**	**0.79**	**0.16**	**0.45**	**0.61**	**1.45**

Source: **compiled by the authors**

Table 2. **Experimental results of the proposed method and peer methods (continuation)**

Method	Dataset											
	Dolls				Reindeer				Laundry			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
AR [26]	0.22	0.35	0.52	0.82	0.23	0.42	0.62	1.12	0.21	0.35	0.52	1.14
Guided [27]	0.29	0.36	0.57	1.15	0.43	0.56	0.89	1.82	0.40	0.54	0.98	1.92
JBU [28]	0.22	0.40	0.76	1.48	0.28	0.52	1.02	1.90	0.27	0.51	0.95	1.97
TGV [29]	0.23	0.35	0.72	2.21	0.33	0.51	1.05	3.07	0.32	0.57	0.24	3.39
Bicubic [30]	0.22	0.38	0.68	1.20	0.32	0.57	1.01	1.89	0.30	0.56	1.06	1.97
MLS [31]	0.25	0.38	0.62	0.99	0.34	0.65	0.77	1.45	0.25	0.41	0.83	1.56
NLMR [33]	0.18	0.33	0.58	1.07	0.22	0.39	0.65	1.30	0.19	0.35	0.57	1.16
Proposed	**0.11**	**0.24**	**0.44**	**0.77**	**0.13**	**0.29**	**0.51**	**1.01**	**0.12**	**0.26**	**0.45**	**0.95**

Source: **compiled by the authors**



a b c d

Fig. 1. **Visual comparison of depth maps:**
a – source image; b – TGV method; c – JGF method; d – proposed method
Source: **compiled by the authors**

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

CONCLUSIONS

This study presents a new method for refining depth maps using color image data. The proposed approach improves the accuracy of depth maps and preserves structural details by employing a structural affinity calculation method based on minimum spanning trees and forests.

The method effectively eliminates edge-leaking artifacts, typical for global minimum spanning trees, through the application of super-pixel segmentation, allowing depth edges to be preserved with minimal distortion.

The integration of an edge inconsistency measurement model into the spanning tree construction process successfully reduces texture-copying artifacts, making the refined depth maps more realistic and accurate.

The dynamic selection of the regularization term's window bandwidth, based on local depth variations, significantly enhances the method's ability to handle different noise levels and depth discontinuities. Experimental results demonstrate that the proposed method achieves the lowest mean absolute deviation across various datasets and upsampling factors. Visual comparisons confirm the superiority of the refined depth maps in preserving details without introducing artifacts.

By utilizing a divide-and-conquer algorithm for spanning tree computations, the method maintains high computational efficiency, making it suitable for practical applications in areas such as computer vision, robotics, and augmented reality. This study represents a significant step forward in depth map refinement, providing robust results across various scenarios while preserving computational efficiency.

Future work is expected to explore the method’s application in real-time systems and its integration with other depth-sensing technologies.

REFERENCES

1. Chang, Y. -J. Kim, S. & Ho, Y. -S. “Depth upsampling methods for high resolution depth map”. *International Conference on Electronics, Information, and Communication (ICEIC)*. Honolulu, HI, USA. 2018. p. 1–4. DOI: https://doi.org/10.23919/ELINFOCOM.2018.8330567.

2. Liu, W., Chen, X., Yang J. & Wu, Q. “Robust color guided depth map restoration”. *In IEEE Transactions on Image Processing*. 2017; 26 (1): 315–327. DOI: https://doi.org/10.1109/TIP.2016.2612826.

3. Hou, J., Ji, Z., Yang, J., Wang, C. & Zheng, F. “MCD-Net: Toward RGB-D Video Inpainting in Real-World Scenes”. *In IEEE Transactions on Image Processing*. 2024; 33: 1095–1108. DOI: https://doi.org/10.1109/TIP.2024.3358675.

4. Jia, Z., Jin, Q., Ng, M. K. & Zhao, X. -L. “Non-local robust quaternion matrix completion for large-scale color image and video inpainting”. *In IEEE Transactions on Image Processing*. 2022; 31: 3868–3883. DOI: https://doi.org/10.1109/TIP.2022.3176133.

5. Xu, R., Li, X., Zhou, B. & Loy, C. C. “Deep flow-guided video inpainting”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Long Beach, CA, USA. 2019. p. 3718–3727. DOI: https://doi.org/10.1109/CVPR.2019.00384.

6. Wang, W., Liu, F., Hao, J. & Shi, C. “An occlusion removal approach for surgical instruments based on the optical flow-guided models”. *IEEE International Conference on Mechatronics and Automation (ICMA)*. Tianjin, China. 2024. p. 1532–1537. DOI: https://doi.org/10.1109/ICMA61710.2024.10632993.

7. Zhang, Y. & Funkhouser, T. “Deep depth completion of a single RGB-D image”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA. 2018. p. 175–185. DOI: https://doi.org/10.1109/CVPR.2018.00026.

8. Wang, J., et al. “MSF-Net: Multi-Scale Feedback reconstruction for guided depth map super-resolution”. In: *IEEE Transactions on Circuits and Systems for Video Technology*. 2024; 34 (2): 709–723. DOI: https://doi.org/10.1109/TCSVT.2023.3288339.

9. Ferstl, D., Reinbacher, C., Ranftl, R., Ruether, M. & Bischof, H. “Image guided depth upsampling using anisotropic total generalized variation”. *IEEE International Conference on Computer Vision*. Sydney. NSW, Australia. 2013. p. 993–1000. DOI: https://doi.org/10.1109/ICCV.2013.127.

10. Hui, T.-W. & Ngan, K. N. “Dense depth map generation using sparse depth data from normal flow”. *IEEE International Conference on Image Processing (ICIP)*. Paris, France. 2014. p. 3837–3841. DOI: https://doi.org/10.1109/ICIP.2014.7025779.

11. Pourazad, M. T., Zhou, D., Lee, K., Karimifard, S., Ganelin, I. & Nasiopoulos, P. “Improving depth map compression using a 3-phase depth map correction approach”. *IEEE International Conference on*

Multimedia & Expo Workshops (ICMEW). Turin, Italy. 2015. p. 1–6. DOI: https://doi.org/10.1109/ICMEW.2015.7169790.

12. Garcia, F., Aouada, D., Solignac, T., Mirbach, B. & Ottersten, B. “Real-time depth enhancement by fusion for RGB-D cameras”. *IET Computer Vision.* 2013; 5 (5): 293–413. DOI: https://doi.org/10.1049/iet-cvi.2012.0289.

13. Banitalebi-Dehkordi, A., Pourazad, M. T. & Nasiopoulos, P. “A study on the relationship between depth map quality and the overall 3D video quality of experience”. *3DTV Vision Beyond Depth (3DTV-CON)*. Aberdeen, UK. 2013. p. 1–4. DOI: https://doi.org/10.1 109/3DTV.2013.6676650.

14. Xu, D., Fan, X., Zhang, S., Wang, Y., Zhao, D. & Gao, W. “Multiscale directional fusion for depth map super resolution with denoising”. *ICASSP-IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brighton, UK. 2019. p. 2342–2346, DOI: https://doi.org/10.1109/ICASSP.2019.8683638.

15. Hambarde, P., Dudhane, A., Patil, P. W., Murala, S. & Dhall, A. “Depth estimation from single image and semantic prior”. *IEEE International Conference on Image Processing (ICIP)*. Abu Dhabi, United Arab Emirates. 2020. p. 1441–1445. DOI: https://doi.org/10.1109/ICIP40778.2020.9190985.

16. Yang, X., Chang, Q., Liu, X., He S. & Cui, Y. “Monocular depth estimation based on multi-scale depth map fusion”. *IEEE Access.* 2021; 9: 67696–67705. DOI: https://doi.org/10.1109/ACCESS.2021.3076346.

17. Lin, K. -Y. & Hang, H. -M. “Depth map enhancement on RGB-D video captured by kinect V2”. *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. Honolulu, HI, USA. 2018. p. 1530–1535. DOI: https://doi.org/10.23919/APSIPA.2018.8659552.

18. De Silva, D. V. S. X., Fernando, W. A. C., Worrall, S. T. & Kondoz, A. M. “A novel depth map quality metric and its usage in depth map coding”. *3DTV Conference: The True Vision – Capture, Transmission and Display of 3D Video (3DTV-CON)*. Antalya, Turkey. 2011. p. 1–4. DOI: https://doi.org/10.1109/3DTV.2011.5877203.

19. Hui Ding, Zhaohui Li & Dongmei Li. “Depth map pre-processing algorithm for compression based on 3D-HEVC scheme”. *IEEE 16th International Conference on Communication Technology (ICCT)*. Hangzhou. 2015. p. 290–294. DOI: https://doi.org/10.1109/ICCT.2015.7399845.

20. Zhang, S., Wang, C. & Chan, S. C. “A new high resolution depth map estimation system using stereo vision and depth sensing device”. *IEEE 9th International Colloquium on Signal Processing and its Applications.* Kuala Lumpur, Malaysia. 2013. p. 49–53, DOI: https://doi.org/10.1109/CSPA.2013.6530012.

21. Yeo, Y.-J., Sagong, M.-C., Shin, Y. -G., Jung, S. -W. & Ko, S.-J. “Simple yet effective way for improving the performance of depth map super-resolution”. *IEEE Signal Processing Letters*. 2020; 27: 2099–2103. DOI: https://doi.org/10.1109/LSP.2020.3039429.

22. Kondratyev, S. B., Antoshchuk, S. G. & Hodovychenko, M. A. “Methods for refining the depth map obtained from depth sensors”. *Applied Aspects of Information Technology*. 2024; 7 (4): 336–347. DOI: https://doi.org/10.15276/aait.07.2024.23.

23. Bevilacqua, M., Aujol, J. -F., Brédif, M. & Bugeau, A.”Visibility estimation and joint inpainting of lidar depth maps”. *IEEE International Conference on Image Processing (ICIP)*. Phoenix, AZ, USA. 2016. p. 3503–3507. DOI: https://doi.org/10.1109/ICIP.2016.7533011.

24. Andorko, I., Corcoran, P. & Bigioi, P. “Depth map generation and depth layer separation for information customization in computer gaming”. *2nd International IEEE Consumer Electronics Society's Games Innovations Conference.* Hong Kong, China. 2010. p. 1–4. DOI: https://doi.org/10.1109/ICEGIC.2010.5716888.

25. “Middlebury Stereo Datasets”. – Available from: https://vision.middlebury.edu/stereo/data. – [Accessed: Nov, 2023].

26. Gabdullin, B., et al. “DepthART: Monocular Depth Estimation as Autoregressive Refinement Task”. *arXiv:2409.15010*. 2024. DOI: https://doi.org/10.48550/arXiv.2409.15010.

27. Ochotorena, C. A., Ochotorena, C. N. & Dadios, E. “Gradient-guided filtering of depth maps using deep neural networks”. *International Conference on Humanoid, Nanotechnology, Information Technology,Communication and Control, Environment and Management (HNICEM)*. Cebu, Philippines. 2015. p. 1–8. DOI: https://doi.org/10.1109/HNICEM.2015.7393265.

28. Kim, J., Jeon, G. & Jeong, J. “Joint-Adaptive bilateral depth map upsampling”. *Signal Processing: Image Communication.* 2014; 29 (4): 506–513. DOI: https://doi.org/10.1016/j.image.2014.01.011.

 ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

29. Kim, S.-Y., Lee, E.-K. & Ho, Y.-S. “Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera”. *In IEEE Transactions on Broadcasting*. 2008; 54 (4): 732–740. DOI: https://doi.org/10.1109/TBC.2008.2002338.

30. Liu, W., Chen, X., Shen, C., Yu, J., Wu, Q. & Yang, J. “Robust Guided Image Filtering”. *arXiv:1703.09379*. 2017. p. 1–14. DOI: https://doi.org/10.48550/arXiv.1703.09379.

31. Bose, N. K. & Ahuja, N. A. “Superresolution and noise filtering using moving least squares”. In: *IEEE Transactions on Image Processing*. 2006; 15 (8): 2239–2248. DOI: https://doi.org/10.1109/TIP.2006.877406.

32. Diebel, J. & Thrun, S. “An application of Markov random fields to range sensing”. *In Proceedings of the 18th International Conference on Neural Information Processing Systems (NIPS'05). MIT Press*. Cambridge, MA, USA. 2005. p. 291–298.

33. Park, J., Kim, H., Tai, Y.-W., Brown, M. S. & Kweon, I. S. “High-Quality depth map upsampling and completion for RGB-D cameras”. *In IEEE Transactions on Image Processing*. 2014; 23 (12): 5559–5572. DOI: https://doi.org/10.1109/TIP.2014.2361034.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding this study, including financial, personal, authorship or other, which could influence the research and its results presented in this article

Received 10.09.2024
Received after revision 14.11.2024
Accepted 19.11.2024

DOI: https://doi.org/10.15276/hait.07.2024.25

УДК 004.93

Надійне уточнення карти глибин за допомогою кольорового зображення

Кондратьєв Сергій Борисович[1)]
ORCID: https://orcid.org/0000-0003-4975-5757; voshodvostok@gmail.com
Антощук Світлана Григорівна[1)]
ORCID: https://orcid.org/0000-0002-9346-145X; asg@opu.ua. Scopus Author ID: 8393582500
Годовиченко Микола Анатолійович[1)]
ORCID: https://orcid.org/0000-0001-5422-3048; hodovychenko@op.edu.ua. Scopus Author ID:57188700773
Устенко Сергій Анатолійович[1)]
ORCID: https://orcid.org/0000-0003-4968-1233; ustenko.s.a@gmail.com. Scopus Author ID: 57207577774
[1)] Національний університет «Одеська Політехніка», пр. Шевченка, 1. Одеса, Україна, 65044

АНОТАЦІЯ

Карти глибини є важливим інструментом для багатьох застосувань, оскільки вони надають просторову інформацію про розташування об’єктів у сцені. Вони відіграють ключову роль у таких сферах, як комп’ютерний зір, робототехніка, доповнена та віртуальна реальність, автономні системи та медична візуалізація. Однак створення точних і високоякісних карт глибини залишається складним завданням через такі проблеми, як артефакти копіювання текстур, протікання границь та спотворення границь глибини. У цьому дослідженні запропоновано новий метод уточнення карт глибини, що інтегрує інформацію з кольорових зображень, поєднуючи структурні та статистичні підходи для досягнення високих результатів. Запропонований підхід застосовує структурний метод для обчислення афінностей у рамках регуляризації, використовуючи мінімальні остовні дерева (MST) та мінімальні остовні ліси (MSF). Для запобігання побудови MST через границі глибини використовується сегментація на суперпікселі, що дозволяє уникнути протікання границь і зберігати деталі. Модель оцінки несумісності границь додатково зменшує артефакти копіювання текстур. Крім того, адаптивне регуляризаційне вікно динамічно налаштовує свою ширину залежно від локальних змін глибини, що дозволяє ефективно справлятися з шумами та зберігати чіткість границь глибини. Експериментальна оцінка на різних наборах даних демонструє високу точність і надійність методу. Він стабільно забезпечує найнижче середнє абсолютне відхилення (MAD) у порівнянні з існуючими техніками при різних коефіцієнтах масштабування, включаючи 2×, 4×, 8× та 16×. Візуальні оцінки підтверджують здатність методу створювати карти глибини без артефактів копіювання текстур і розмиття границь, максимально наближені до еталонних даних. Обчислювальна ефективність забезпечується завдяки алгоритму «розділяй і владарюй» для обчислення остовних дерев, що зменшує складність при збереженні точності. Дослідження підкреслює важливість поєднання структурної та статистичної інформації у вдосконаленні карт глибини. Подолавши обмеження існуючих методів, запропонований підхід надає практичне рішення для покращення карт глибини у додатках, що потребують високої точності та ефективності, таких як робототехніка, віртуальна реальність та автономні системи. У майбутньому передбачається дослідження застосування методу в реальному часі та його інтеграція з передовими технологіями отримання глибини.

Ключові слова: карти глибин; 3D-реконструкція; обробка зображень; просторовий аналіз даних; уточнення даних; сенсорна візуалізація; виявлення країв; зменшення шуму; вимірювання глибини; обчислювальна візуалізація; доповнена реальність; автономні системи

ABOUT THE AUTHORS

Sergey B. Kondratyev - Senior lecturer, Department of Artificial Intelligence and Data Analysis. Odesa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0003-4975-57574; voshodvostok@gmail.com
Research field: Video processing; motion tracking; stereoscopic vision; depth map; emotion intellect

Кондратьєв Сергій Борисович - ст. викладач кафедри Штучного інтелекту та аналізу даних. Одеський національний політехнічний університет, пр. Шевченка, 1. Одеса, 65044,Україна

Svitlana G. Antoshchuk - Doctor of Engineering Sciences, Professor, Head of Computer Systems Institute. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-9346-145X;asg@op.edu.ua. Scopus Author ID: 8393582500
Research field: Pattern recognition; deep learning;object tracking;face recognition; graphic images formation and processing

Антощук Світлана Григорівна - доктор технічних наук, професор, директор Інституту комп'ютерних систем Національний університет «Одеська Політехніка», пр. Шевченка, 1. Одеса, 65044,Україна

Mykola A. Hodovychenko – PhD (Eng.), Associate professor, Department of Artificial Intelligence and Data Analysis. Odesa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0001-5422-3048; hodovychenko@od.edu.ua. Scopus Author ID: 57188700773
Research field: Deep learning; data mining; smart cities; video processing; motion tracking; project-based learning; patter recognition

Годовиченко Микола Анатолійович - кандидат технічних наук, доцент Інституту штучного інтелекту та робототехніки, кафедри Штучного інтелекту та аналізу даних. Одеський національний політехнічний університет, пр. Шевченка, 1. Одеса, 65044,Україна

Serhii A. Ustenko - Doctor of Physical and Mathematical Sciences, Associate professor, Department of Artificial Intelligence and Data Analysis. Odesa National Polytechnic University, 1, Shevchenko Ave. Odesa, 65044, Ukraine
ORCID: https://orcid.org/0000-0003-4968-1233; ustenko.s.a@gmail.com. Scopus Author ID: 57207577774
Research field: Deep learning; data mining; smart cities; video processing; motion tracking; project-based learning; patter recognition

Устенко Сергій Анатолійович - доктор фізико-математичних наук, доцент Інституту штучного інтелекту та робототехніки, кафедри Штучного інтелекту та аналізу даних. Одеський національний політехнічний університет, пр. Шевченка, 1. Одеса, 65044,Україна