DOI: https://doi.org/10.15276/aait.07.2024.23
UDC 004.93

Methods for refining the depth map obtained from depth sensors

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. Scopus Author ID: 57188700773
[1)] Odesa Polytechnic National University, 1, Shevchenko Avenue. Odesa, 65044, Ukraine

ABSTRACT

Depth maps are essential in applications such as robotics, augmented reality, autonomous vehicles, and medical imaging, providing critical spatial information. However, depth maps from sensors like time-of-flight (ToF) and structured light systems often suffer from low resolution, noise, and missing data. Addressing these challenges, this study presents an innovative method to refine depth maps by integrating high-resolution color images. The proposed approach employs both hard- and soft-decision pixel assignment strategies to adaptively enhance depth map quality. The hard-decision model simplifies edge classification, while the soft-decision model, integrated within a Markov Random Field framework, improves edge consistency and reduces noise. By analyzing discrepancies between edges in depth maps and color images, the method effectively mitigates artifacts such as texture-copying and blurred edges, ensuring better alignment between the datasets. Key innovations include the use of the Canny edge detection operator to identify and categorize edge inconsistencies and anisotropic affinity calculations for precise structural representation. The soft-decision model introduces advanced noise reduction techniques, improving depth map resolution and preserving edge details better than traditional methods. Experimental validation on Middlebury benchmark datasets demonstrates that the proposed method outperforms existing techniques in reducing Mean Absolute Difference values, especially in high-upscaling scenarios. Visual comparisons highlight its ability to suppress artifacts and enhance edge sharpness, confirming its effectiveness across various conditions. This approach holds significant potential for applications requiring high-quality depth maps, including robotics, augmented reality, autonomous systems, and medical imaging. By addressing critical limitations of current methods, the study offers a robust, versatile solution for depth map refinement, with opportunities for real-time optimization in dynamic environments.

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. "Methods for refining the depth map obtained from depth sensors". *Applied Aspects of Information Technology.* 2024; Vol. 7 No. 4: 336–347. DOI: https://doi.org/10.15276/aait.07.2024.23**

INTRODUCTION, FORMULATION OF THE PROBLEM

Depth maps are a cornerstone of modern technology, providing spatial information critical for understanding and interacting with three-dimensional environments. Their importance spans a diverse range of applications, from robotics and automation to advanced imaging and modeling [1].
In robotics, depth maps enable machines to perceive their surroundings with precision, facilitating tasks such as navigation, object recognition, and manipulation in dynamic environments. This capability is fundamental for the development of autonomous systems and industrial automation, where spatial awareness is paramount.

In augmented and virtual reality, depth maps enhance user experience by accurately blending virtual elements with real-world environments. They allow systems to perceive spatial relationships, creating immersive and interactive virtual spaces that transform entertainment, education, and training. Similarly, autonomous vehicles depend on depth maps for safety and functionality [2].

They provide the spatial data necessary for obstacle detection, scene interpretation, and path planning, ensuring reliable operation in complex and variable environments.

Beyond navigation and interactivity, depth maps are crucial in three-dimensional reconstruction and modeling. They enable the creation of precise models of objects and environments, serving industries such as gaming, architecture, and cultural heritage preservation.

These models facilitate realistic renderings, design optimization, and the digital preservation of historically significant sites and artifacts. In medical imaging, depth maps enhance visualization

ISSN 2617-4316 (Print)
ISSN 2663-7723 (Online)

techniques and enable the reconstruction of anatomical structures, supporting accurate diagnostics and surgical planning.

Depth maps also play a vital role in surveillance and security systems, aiding in object tracking and identification under challenging conditions such as low light or crowded scenes. Their ability to provide depth information adds an extra layer of reliability to systems requiring robust performance in real-world settings.

Modern methods for generating depth maps can be broadly categorized into passive and active approaches. Passive methods rely on analyzing two- or multi-view color images using stereo correspondence algorithms. These techniques, extensively studied over the past decades, estimate depth values based on either local or global image processing [3].

Local methods are characterized by higher computational speed due to their independent processing of each pixel, but they often exhibit lower accuracy compared to global approaches, which optimize depth values for the entire scene simultaneously. The primary limitations of passive methods include difficulties in handling low-textured regions and challenges arising from occlusions.

In contrast, active methods employ specialized depth sensors that enable depth map generation at frame rates comparable to those of color cameras. The most widely used active depth sensing technologies include time-of-flight (ToF) sensors and structured light sensors (Fig. 1) [4].

ToF sensors calculate depth by measuring the phase shift between emitted and reflected infrared light, but they often produce noisy depth maps with low resolution. Structured light sensors project an infrared pattern onto the scene, which is then analyzed to compute depth, offering higher resolution. However, depth maps obtained through this method frequently suffer from artifacts such as “holes”, caused by occlusions, low surface reflectivity, or distortions in the projected pattern [5].

The common challenges associated with depth maps include low resolution, noise, and missing data in certain areas. These deficiencies can be partially mitigated through enhancement methods that leverage color images, which provide additional structural information about the scene. The strong correlation between texture and depth distribution allows for significant improvements in the quality of depth maps. Consequently, integrating depth data with color imagery represents a promising avenue for advancing depth sensing technologies.

Fig. 1. **Color image and depth map, obtained with structured light sensor**
Source: **compiled by the [4]**

Thus, **the purpose of this study** is to develop a method for depth maps refinement using color images. The proposed approach integrates depth data with color imagery to enhance the quality of depth maps and fill missing regions.

1. LITERATURE REVIEW

Depth map enhancement is a complex task encompassing two primary objectives: increasing spatial resolution (super-resolution) and restoring missing data (depth completion). These objectives are particularly relevant for data obtained using ToF sensors and structured-light sensors.

Super-resolution aims to improve the spatial resolution of depth maps, a critical requirement for applications demanding high-precision visualization. Depth completion, on the other hand, focuses on addressing areas with missing depth data, which are typical in real-world scenarios [6].

Despite their differing goals, these tasks share a common foundation and can be unified under a single formulation. Modern methods, independent of

external datasets, leverage the principle of joint analysis of edges in depth maps and corresponding color images. This approach utilizes the additional information provided by the texture of the color image to refine the structure of the depth map [7].

The methods addressing these challenges are broadly classified into two categories: filter-based methods and optimization-based techniques.

Filter-based methods are efficient due to their localized nature. One of the pioneering works in this area is the joint bilateral upsampling technique proposed in [8]. This method uses high-resolution color images to refine the edges of low-resolution depth maps through bilateral filtering techniques.

Subsequently, the approach was improved in [9], where the traditional separation of color space and spatial distances was replaced with a unified geodesic space, leading to enhanced results. Another work [10] proposed an iterative method based on cost volumes, which involves multiple depth candidates. Each candidate is refined using joint bilateral upsampling.

An important contribution was made in [11], where the authors introduced guided filtering, which models a linear relationship between the output image and the guiding image. This method assumes that edges in the output are present only when edges exist in the input.

Further advancements were achieved in [12], which introduced a weighted mode filtering method based on a joint histogram of depth candidates. This method minimizes the L1 norm, making it more robust to outliers compared to L2 norm minimization. The authors of [13] proposed trilateral filtering, which incorporates local gradient information of depth maps, providing a significant advantage over bilateral upsampling.

The method proposed in [14] employs an onion-peeling filtering procedure to consider local depth gradients for super-resolution. However, all filter-based methods are limited in their ability to suppress noise due to the localized nature of their solutions.

Optimization-based approaches, in contrast to filter-based methods, are more robust to noise and can model more complex dependencies. One of the earliest works in this domain was presented in [15], where depth map super-resolution was formulated as a multi-label optimization problem, solved using Markov Random Fields.

The approach was refined in [16], where a more adaptive data term tailored to the characteristics of depth maps was introduced. Another work [17] proposed dynamic Markov Random Fields, extending traditional spatial Markov Random Fields by incorporating temporal information. This significantly improved the accuracy and robustness of super-resolution in dynamic scenes.

Further advancements were introduced in [18], which included non-local regularization using edge, gradient, and segmentation information extracted from high-resolution color images.

An innovative contribution was made in [19] where the authors applied generalized second-order smoothness constraints guided by an anisotropic diffusion tensor derived from color images.

The authors of [20] proposed an auto-regressive model that creates a predictor for each pixel based on local correlations in the initial depth map and non-local similarities in registered high-quality color images. In [21], a robust M-estimator-based regularization term was developed, enabling the method to account for inconsistencies between depth maps and color images.

Thus, the quality of edges in low-resolution depth maps can be significantly enhanced by leveraging additional information from the corresponding edges in color images. This approach is based on the assumption of a strong correlation between the edge structures in the color image and the depth map. However, this assumption does not always hold true, as substantial discrepancies between the two can often be observed [22].

The improper utilization of guiding information provided by the color image may result in two primary issues: texture-copying artifacts and blurred edges on the depth map.

Texture-copying artifacts occur when inherently smooth regions in the depth map are misinterpreted as textured due to the presence of corresponding textures in the color image. Conversely, blurred edges typically arise when relatively homogeneous regions in the color image align with areas in the depth map that exhibit strong gradients [23].

Previously proposed methods aimed at improving depth maps have sought to balance the contributions of the original depth map and the corresponding color image. However, these methods have significant limitations. The primary shortcoming is their inability to explicitly evaluate edge inconsistencies between the depth map and the color image. This lack of evaluation restricts the ability to adaptively regulate the influence of guiding information from the color image during the depth map refinement process [24].

To address these limitations, an advanced depth map enhancement method has been developed, incorporating mechanisms for assessing edge

discrepancies through both hard and soft decision strategies. These approaches enable more precise control over the integration of edge information from the color image, effectively mitigating common errors such as texture copying or edge blurring [25].

The proposed method is grounded in a comprehensive analysis of the relationship between the edge structures of the depth map and the color image. This allows for a substantial improvement in the accuracy and quality of the resulting depth maps. By adopting an adaptive approach that accounts for both local and global scene characteristics, this method represents a significant advancement in the development of more robust and versatile algorithms for depth map enhancement.

2. PROPOSED METHOD

To evaluate edge inconsistency between a low-resolution depth map and a high-resolution color image, it is necessary to consider several key aspects that determine the accuracy and reliability of such an analysis.

First and foremost, for a precise measurement of the inconsistency between the depth edge map and the color edge map, their resolutions must be unified. This requirement arises because discrepancies in resolution make it challenging to accurately align the structures and positions of the edges. In this context, the depth map, which may have lower resolution or missing regions, must be pre-interpolated to match the resolution of the color image. This interpolation can be performed using either regular methods (grid-based) or irregular methods (considering spatial distribution of the data), ensuring uniform resolution before edge detection [26].

A fundamental characteristic underlying this analysis is the structural similarity between the color image and the corresponding depth map. This similarity becomes particularly evident when comparing their binary edge maps, which capture the primary contours and boundaries of objects. The proposed method is based on evaluating the inconsistency between the binary edge maps generated separately from the color image and the depth map.

To detect edges in the coarsely interpolated depth map and the corresponding color image, the Canny edge detection operator is employed. This operator is widely recognized for its ability to extract clear and stable edges, even in the presence of noisy data. However, due to the low resolution or noise in the depth map, the detected edges may shift from their actual positions. This introduces additional challenges when comparing these edges with the registered edges from the high-resolution color image [27].

Based on the analysis of such deviations, inconsistent edges in the depth map can be categorized into two types. The first type includes edges degraded by coarse interpolation, which can be refined using guiding information from high-resolution color edges. The second type represents true inconsistencies, stemming from fundamental differences between the depth map and the color image.

2.1. Hard-decision pixel assignment

An effective approach for classifying these types of inconsistencies is the hard-decision method, which definitively assigns each edge pixel to one of these categories. Inspired by the principles of error correction coding, where the number of errors must remain below a threshold to allow successful correction, the following strategy is proposed: if the displacement between an edge in the depth map and the nearest edge in the color image is below a predefined threshold, the edge is considered degraded due to interpolation. Otherwise, it is classified as a true inconsistency.

This approach not only simplifies the analysis process but also provides a more robust mechanism for managing edge information, enabling effective correction and refinement of the depth map using high-quality color data. Furthermore, the adaptability of this methodology enhances its applicability across a wide range of tasks related to image processing and depth data analysis.

Target function. In accordance with Hammersely-Clifford theorem [28], target function for depth map super-resolution could be calculated as:

$$\begin{aligned} M' &= \arg\min_{m_k \in M} \sum_{v_k \in V} \xi_c^k H_t(m_k, v_k) \\ &+ \xi \sum_k \sum_{l \in N_k} \xi_f^{kl} H_r(m_k, m_l), \end{aligned} \tag{1}$$

$$H_t(m_k, v_k) = |m_k - v_k|, \tag{2}$$

$$H_r(m_k, m_l) = |m_k - m_l|, \tag{3}$$

where M' is depth map; V is perceived values of depth attribute; k, l are pixels of refined depth map; v_k is perceived depth attribute of pixel p; N_k is surrounding pixels of pixel p; H_t represents the data component that reflects the consistency between the

refined depth attribute and the perceived values; H_r represents the regularization component that promotes a piecewise smooth solution and discourages differing depth assignments among adjacent pixels; ξ is utilized to balance the influence of the data component and the regularization component; ξ_c^k is confidence of perceived v_k; ξ_f^{kl} is anisotropic affinity of k, l which incorporating the proposed hard-decision pixel assignment.

Outliers Discovery. Depth values observed for pixels located at depth edges in low-resolution depth maps are unreliable. This unreliability arises due to the blurring effect caused by the mixing of depth values from two distinct layers. Such distortions make these values unsuitable for inclusion in the construction of the data term, as they significantly reduce the accuracy of the model.

In the proposed approach, the Canny edge detection operator is employed to identify edges in the low-resolution depth map. This operator has proven to be an effective tool for detecting edges, even in low-quality input data. By analyzing gradients, it enables the identification of key object boundaries, which, in this context, serve as indicators of unreliable pixels.

To designate pixels located on edges, a binary variable ξ_c^k is introduced. The value of this variable is determined as follows: if a pixel k is classified as an edge pixel, ξ_c^k is assigned a value of 0, indicating its exclusion from the data term. Conversely, if the pixel does not belong to an edge, ξ_c^k is set to 1, allowing the pixel to be included in the data component construction process.

This proposed method effectively excludes pixels with unreliable depth values, minimizing the influence of artifacts and distortions typically associated with blurred edge regions. As a result, it enhances the accuracy of depth map processing and analysis. Furthermore, the use of the Canny operator provides a straightforward yet effective procedure for edge detection, making this approach applicable to a wide range of tasks in image processing and depth data analysis.

Anisotropic Affinity obtaining. For each pixel l situated near pixel k, the anisotropic affinity ξ_f^{kl}, which quantifies the relationship between the pixel pair k, l, is computed based on the values of ξ_f^k and ξ_f^l. These values are influenced by the structural attributes of the pixels within the color image and the coarsely interpolated depth map, and their determination follows the rules outlined below:

1) if pixel k is consistently positioned either on an edge or within a homogeneous region in both the color image and the coarsely interpolated depth map, it is presumed that the color distribution aligns with the depth distribution in the vicinity of k. In such cases, ξ_f^k is derived using a weighting function proposed in the joint bilateral upsampling technique [29], which evaluates the influence of guidance provided by the color image. This condition is referred to as "Color-guided", highlighting the coherence between the two data sources;

2) when pixel k lies on an edge in the color image but not on the coarsely interpolated depth map, a search window is established on the depth map. If edges are found within this search window, ξ_f^k is categorized as "Color-guided". Otherwise, this situation is interpreted as genuine edge inconsistency. Since pixel k resides in a uniform region of the depth map, ξ_f^k is assigned a high value to suppress discrepancies in label assignments between neighboring pixels. This scenario is identified as "Uniform region" signifying the steadiness of depth information in the area;

3) if pixel k is located on an edge in the coarsely interpolated depth map but not in the color image, a search window is defined on the color image. Should edges be present within this window, ξ_c^k is classified as "Color-guided". Otherwise, this condition is regarded as a true edge inconsistency;

4) since pixel k is positioned near depth edges, ξ_c^k is assigned a low value to encourage distinct label assignments for neighboring nodes. This case is referred to as "Edge-adjacent" emphasizing the importance of accommodating local variations in depth.

This methodology for computing pixel affinities takes into account the local traits of depth data and its alignment with color information, facilitating a more precise and flexible model design. Consequently, it ensures a detailed and reliable representation in the processing of depth data for applications in computer vision. As derived from the analysis presented, ξ_f^k and ξ_f^l are calculated as follows:

$$\xi_f^{\{k;l\}} = \begin{cases} e^{-\frac{\Delta L_{kl}^2}{\sigma^2}}, Color-guided \\ e^{-\frac{\Delta L_{sm}^2}{\sigma^2}}, Edge-adjacent\cdot \\ e^{-\frac{\Delta L_{lg}^2}{\sigma^2}}, Uniform\ region, \end{cases} \quad (4)$$

where ΔL_{kl} is the brightness difference between pixels k and l; ΔL_{sm} =1 and ΔL_{lg} = 254.

Due to the symmetrical relationship between pixels k and l in the analyzed pair, the value of ξ_f^{kl}, representing their mutual correlation, is derived based on ξ_f^k and ξ_f^l. Specifically, if ξ_f^k and ξ_f^l yield

the same classification, ξ_f^{kl} can be determined unequivocally, removing any chance of ambiguity. Moreover, according to the definitions of the categories, it is evident that if ξ_f^k is categorized as a "Uniform region", ξ_f^l cannot fall into the «Edge-adjacent» category, and vice versa. This mutual exclusion guarantees consistency in the attributes of neighboring pixels and avoids contradictions during processing.

In more complex scenarios, when ξ_f^k is classified as "Color-guided" and ξ_f^l belongs to a different category, this indicates that pixel l is situated near an edge, either on the color image or on the depth map. Under these circumstances, the value of ξ_f^{kl} should be determined by ξ_f^l, allowing the approach to consider local scene characteristics and maintain precision in edge handling.

For better understanding and analysis, all possible cases are outlined in Table 1. The table's first column and row enumerate the possible sets of values for ξ_f^k and ξ_f^l, respectively.

For each specified combination of ξ_f^k and ξ_f^l, the corresponding value of ξ_f^{kl} is listed within the table. This representation provides a clear framework for determining pixel pair correlations, enhancing the comprehensibility of the methodology and ensuring accuracy in depth data analysis.

Table 1. **Evaluation of ξ_f^{kl} value based on ξ_f^k and ξ_f^l values**

Value of ξ_f^k	Value of ξ_f^l		
	"Color-guided"	"Edge-adjacent"	"Uniform region"
"Color-guided"	"Color-guided"	"Edge-adjacent"	"Uniform region"
"Edge-adjacent"	"Edge-adjacent"	"Edge-adjacent"	---
"Uniform region"	"Uniform region"	---	"Uniform region"

Source: **compiled by the authors**

By implementing a structured process for classifying and defining the values of ξ_f^{kl}, this method offers flexibility in handling various pixel types, making it a robust and adaptable solution for applications involving image processing and depth map evaluation.

2.2. Soft-decision pixel assignment

Given the shortcomings of the edge inconsistency measurement technique based on hard-decision, this section introduces an enhanced approach that utilizes a soft-decision edge inconsistency measurement.

This method allows for a more accurate and quantitative evaluation of the inconsistency between depth edges and their corresponding color edges, which is essential for improving the precision of depth data processing.

The proposed approach is integrated into a Markov Random Field framework, which demonstrates a significantly stronger capability in reducing texture-copying artifacts and maintaining the integrity of depth edges compared to its hard-decision counterpart.

A major motivation for adopting this improved method is the inherent limitations of multi-label optimization using graph cut algorithms, which often fail to achieve an exact global minimum. Additionally, depth values are typically stored in millimeters as continuous floating-point values:

$$\begin{aligned} M' &= \arg\min_{m_k \in M} \sum_{v_k \in V} H_t(m_k, v_k) \\ &+ \xi \sum_k \sum_{l \in N_k} \xi_f^{kl} H_r(m_k, m_l), \end{aligned} \tag{5}$$

$$H_t(m_k, v_k) = (m_k - v_k)^2, \tag{6}$$

$$H_r(m_k, m_l) = (m_k - m_l)^2. \tag{7}$$

where ξ_f^{kl} is calculated using the suggested soft-decision edge inconsistency measurement.

The key differences between the proposed model and the one outlined in the previous section are as follows:

1) assuming Gaussian noise in the original low-resolution depth map, the data component and regularization component are formulated using a quadratic function rather than absolute values. This modification results in improved noise reduction performance and significantly enhances the overall quality of depth map refinement;

2) unlike the hard-decision method, which suffers from limited accuracy in edge classification, the proposed approach employs a soft-decision measurement. The affinities used in the regularization term are calculated using this soft-decision edge inconsistency measurement, allowing for adaptive consideration of local edge variations and a more precise modeling of the interplay between depth and color data. A detailed

explanation of this method and its implementation is provided in the subsequent subsections.

This innovative approach greatly enhances the scope of depth map analysis and processing, ensuring more accurate artifact suppression and superior preservation of structural scene details. It represents a significant step forward in developing adaptive methodologies capable of effectively handling diverse types of depth data in computer vision and image processing tasks.

The assessment of discrepancies between the edge map of a color image and the edge map of a depth map can be framed as a bidirectional evaluation of edge map quality.

However, in previous researches, the conventional method for analyzing the quality of edge maps involves measuring the positional offset of each edge pixel relative to its location on a reference map. The scenario under consideration here differs significantly.

In the context of depth map refinement, edge pixels on the depth edge map and their corresponding pixels on the color edge map, which theoretically should coincide in position, often display deviations. These deviations may result from pre-processing operations, such as coarse interpolation (as discussed previously), or from noise introduced by depth sensors.

Consequently, assessing inconsistency based on positional differences of paired edge pixels, as employed in traditional edge quality evaluation methods [30], is not feasible in this case.

Instead, the proposed approach to measuring edge inconsistency focuses on analyzing the structural similarity of edge maps.

This methodology considers not only the local structure formed by neighboring regions around each pixel but also the global arrangement of the entire edge map.

This enables a more precise evaluation of edge alignment in scenarios where simple positional comparisons of edge pixels are insufficient for accurate analysis.

To simplify the explanation, the method is described in terms of a reference edge map and a target edge map. For every edge pixel on the reference map, the approach identifies the best match on the target map within a specified neighborhood around the corresponding position.

This implies that if the edges in the color map and the depth map are well-aligned, the displacement of matched edge pixels will remain within a small range.

Additionally, it is important to account for not only the magnitude of the displacement but also its direction and the uniformity of displacements among all matched edge pixels in a local region [31].

As such, the proposed approach not only captures a more accurate representation of the alignment between edge maps but also incorporates both local and global structural context.

This makes it a more robust solution for analyzing and enhancing depth data in the presence of noise or pre-processing artifacts.

3. EXPERIMENTAL RESULTS

For the experimental evaluation of the proposed method, datasets from the Middlebury benchmark were used [32]. As an additional challenge, a degradation model based on the downsampling method was applied.

The Mean Absolute Difference (MAD) metric was employed as a measure to evaluate the accuracy of the constructed depth map. Mean Absolute Difference refers to a statistical measure that calculates the average of the absolute differences between corresponding elements in two datasets. It is often used to quantify the overall error or deviation between predicted and observed values or between two images in image processing.

Mathematically, it is expressed as:

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

where N is the number of elements in the datasets; x_i and y_i are the corresponding elements from the two datasets.

The proposed method demonstrates the lowest Mean Absolute Difference in most scenarios.

For the challenging case of 16× super-resolution, where the coarsely upsampled depth map introduces substantial errors affecting the quality of the depth edge map, the proposed method achieves the best results in 3 out of 4 cases.

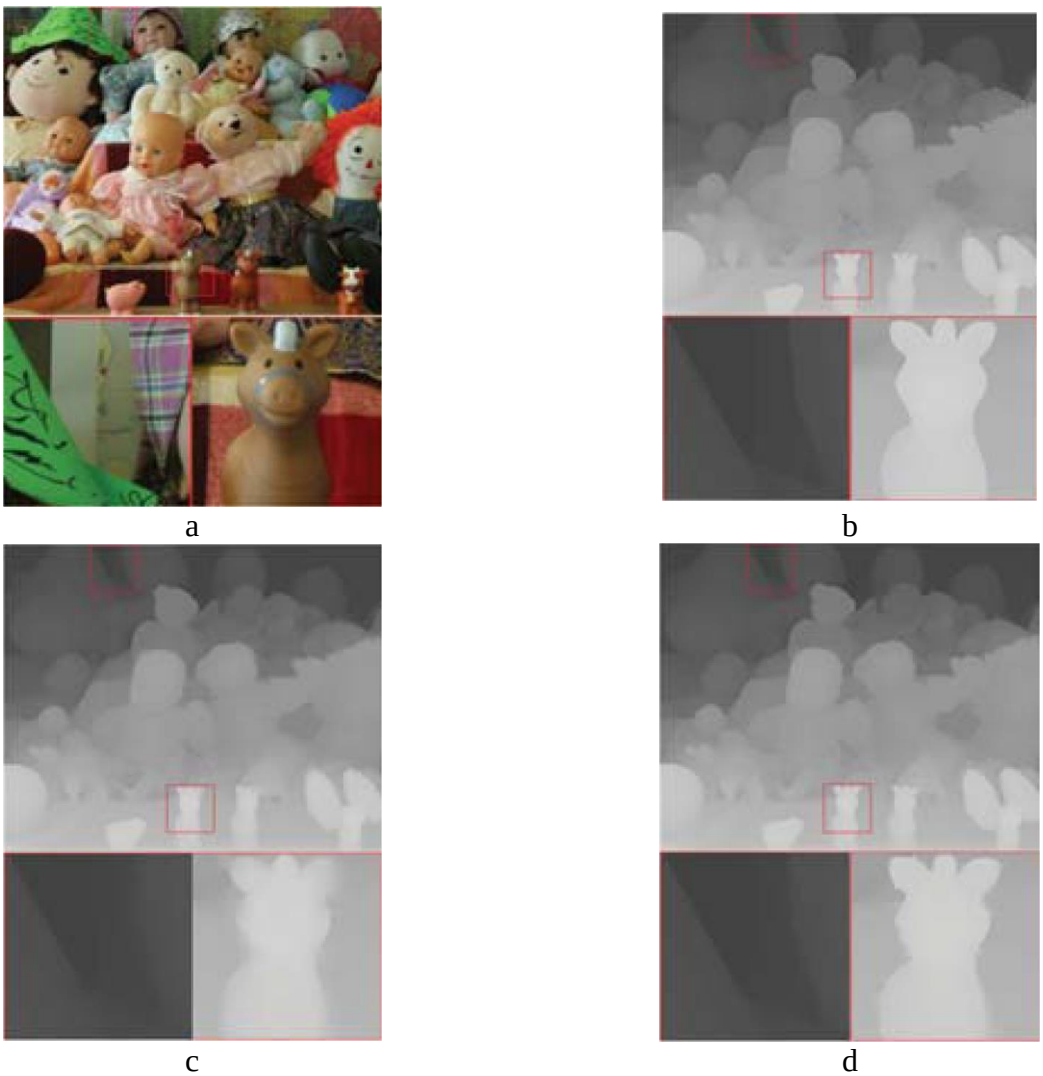
This highlights the robustness of the proposed method to variations in the quality of the depth edge map.

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

Method	Dataset															
	Book				Moebius				Dolls				Reindeer			
	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x	2x	4x	8x	16x
AR	0.13	0.21	**0.36**	0.78	0.13	0.23	0.42	0.83	0.22	0.35	0.52	**0.82**	0.23	0.42	0.62	1.12
Guided	0.23	0.36	0.59	1.15	0.24	0.39	0.61	1.17	0.29	0.36	0.57	1.15	0.43	0.56	0.89	1.82
JBU	0.18	0.37	0.75	1.58	0.19	0.38	0.77	1.48	0.22	0.40	0.76	1.48	0.28	0.52	1.02	1.90
TGV	0.20	0.28	0.43	0.84	0.21	0.30	0.51	0.89	0.23	0.35	0.72	2.21	0.33	0.51	1.05	3.07
Bicubic	0.14	0.30	0.61	1.16	0.14	0.32	0.61	1.15	0.22	0.38	0.68	1.20	0.32	0.57	1.01	1.89
MLS	0.17	0.28	0.47	1.18	0.16	0.26	0.51	0.94	0.25	0.38	0.62	0.99	0.34	0.65	0.77	1.45
Method-H	0.15	0.28	0.49	0.93	0.17	0.32	0.64	1.19	0.19	0.38	0.74	1.45	0.23	0.42	0.77	1.52
Method-S	**0.10**	**0.20**	0.37	**0.74**	**0.11**	**0.21**	**0.39**	**0.81**	**0.12**	**0.26**	**0.49**	0.83	**0.14**	**0.31**	**0.56**	**1.10**

Source: **compiled by the authors**

Fig. 2 presents a visual comparison of depth maps for 8× upscaling. From the visual analysis of the highlighted area, it can be concluded that the AR method suffers from texture-copying artifacts and blurred edges, whereas the proposed method significantly reduces such artifacts.



a b c d

Fig. 2. **Visual comparison of depth maps:**
a – source image; b – real depth map; c – method AR; d – proposed method
Source: **compiled by the authors**

CONCLUSIONS

The study presented a novel approach for refining depth maps obtained from depth sensors, addressing common challenges such as noise, low resolution, and missing data. Through a comprehensive analysis and experimental evaluation, several key conclusions were drawn.

The proposed method effectively leverages high-resolution color images to enhance the quality of depth maps. By integrating depth and color information, the method mitigates artifacts such as texture-copying and blurred edges, ensuring improved depth map fidelity.

A significant innovation of the method is its ability to evaluate edge inconsistencies between depth maps and corresponding color images. By employing both hard- and soft-decision pixel assignment strategies, the approach adapts to local and global scene characteristics, resulting in more accurate and robust depth map refinement.

The soft-decision model, integrated within a Markov Random Field framework, demonstrated superior performance in preserving structural details and suppressing artifacts compared to traditional hard-decision methods. This advancement is crucial for maintaining the integrity of depth edges in noisy environments.

Extensive experiments using datasets from the Middlebury benchmark confirmed the method's efficacy. The proposed approach consistently outperformed peer methods in reducing Mean Absolute Difference, particularly in challenging scenarios requiring significant upscaling of depth maps.

The refinement techniques introduced in this study are applicable across diverse domains, including robotics, augmented reality, autonomous systems, and medical imaging. By enhancing the precision and reliability of depth maps, the proposed method contributes to advancing technologies dependent on accurate three-dimensional spatial data.

In conclusion, the research addresses critical limitations of existing depth sensing and refinement techniques, offering a robust and versatile solution for depth map enhancement. Future work may explore further optimization and real-time applications in dynamic environments.

REFERENCES

1. Ma, F. & Karaman, S. “Sparse-to-Dense: Depth prediction from sparse depth samples and a single image”. *IEEE International Conference on Robotics and Automation (ICRA)*. 2017. DOI: https://doi.org/10.48550/arXiv.1709.07492.

2. Cheng, X., Wang, P. & Yang, R. “Learning depth with convolutional spatial propagation network”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2021; 42 (10): 2361–2379. DOI: https://doi.org/10.1109/TPAMI.2019.2947374.

3. Eldesokey, A., Felsberg, M., & Khan, F. “Confidence propagation through CNNs for guided sparse depth regression”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2020; 42 (10): 2423–2436. DOI: https://doi.org/ 10.1109/TPAMI.2019.2929170.

4. Qiu, J., Cui, Z., Zhang, Y., et al. “DeepLiDAR: Deep surface normal guided depth prediction for outdoor scene from sparse LiDAR data and single color image”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019. pp. 3313–3322. DOI: https://doi.org/10.1109/CVPR.2019.00343.

5. Van Gansbeke, W., Neven, D., Brabandere, B., & Van Gool, L. “Sparse and noisy LiDAR completion with RGB guidance and uncertainty”. *International Conference on Machine Vision Applications (MVA)*. 2019. p. 1–6. DOI: https://doi.org/10.48550/arXiv.1902.05356.

6. Jaritz, M., Charette, R., Wirbel, E., Perrotton, X. & Nashashibi, F. “Sparse and dense data with CNNs: Depth completion and semantic segmentation”. *International Conference on 3D Vision (3DV)*. 2018. p. 52–60. DOI: https://doi.org/10.48550/arXiv.1808.00769.

7. Ma, F., Cavalheiro, G. V. & Karaman, S. " Self-supervised sparse-to-dense: Self-supervised depth completion from LiDAR and monocular camera." *International Conference on Robotics and Automation (ICRA)*. 2019; 29: 3288–3295. DOI: https://doi.org/10.1109/ICRA.2019.8793637.

8. Eldesokey, A., Felsberg, M., Holmquist, K. & Persson, M. “Uncertainty-Aware CNNs for depth completion: Uncertainty from beginning to end”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2020. p. 12011–12020. DOI: https://doi.org/10.1109/CVPR42600.2020.01203.

9. Tang, J., Tian, F. -P., Feng, W., Li, J. & Tan, P. “Learning guided convolutional network for depth completion”. *IEEE Transactions on Image Processing*. 2021; 30: 1116–1129. DOI: https://doi.org/10.1109/TIP.2020.3040528.

10. Wang, Z., Ye, X. & Sun, B., et al. “Depth upsampling based on deep edge-aware learning”. *Pattern Recognition*, 2020; 103: 107274. DOI: https://doi.org/10.1016/j.patcog.2020.107274.

11. Cheng, X., Wang, P., Guan, C., & Yang, R. “CSPN++: Learning Context and Resource Aware Convolutional Spatial Propagation Networks for Depth Completion”. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020; 34 (07): 10615–10622. DOI: https://doi.org/10.1609/aaai.v34i07.6635.

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; 7: 335-345. DOI: https://doi.org/10.1049/iet-cvi.2012.0289.

13. Luo, A., Li, X., Yang, F., Jiao, Z., Cheng, H. & Lyu, S. “Cascade graph neural networks for RGB-D salient object detection”. *Computer Vision – ECCV*. 2020; 12357: 346–364. DOI: https://doi.org/10.1007/978-3-030-58610-2_21.

14. Huang, Y.-K., Wu, T.-H., Liu, Y.-C. & Hsu, W. “Indoor depth completion with boundary consistency and self-attention”. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 2019; p. 1070–1078. DOI: https://doi.org/10.48550/arXiv.1908.08344.

15. Zhang, Y., He, X., Chen, H. & Ren, C. “Depth map super-resolution via learned nonlocal model and enhanced local regularization”. *Signal Processing*. 2024; 218: 109368. DOI: https://doi.org/10.1016/j.sigpro.2023.109368.

16. Imran, S., Liu, X. & Morris, D. “Depth completion with twin surface extrapolation at occlusion boundaries”. *Computer Vision and Pattern Recognition (CVPR)*. 2021. p. 2583–2592. DOI: https://doi.org/10.1109/CVPR46437.2021.00261.

17. Jiajun, Z., Yue, H., Lai, Y.-K., et al. “Deep edge map guided depth super resolution”. *Signal Processing: Image Communication*. 2021; 90: 116040. DOI: https://doi.org/10.1016/j.image.2020.116040.

18. Seki, A. & Pollefeys, M. “SGM-Nets: Semi-global matching with neural networks for depth estimation”. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017. p. 6640–6649. DOI: https://doi.org/10.1109/CVPR.2017.703.

19. Zeng, K., Zhang, H., Wang, W., Wang, Y. & Mao, J. “Deep stereo network with MRF-based cost aggregation”. *In IEEE Transactions on Circuits and Systems for Video Technology*. 2023; 34 (4): 2426–2438. DOI: https://doi.org/10.1109/TCSVT.2023.3312153.

20. Yang, Q., Li, D., Wang L. & Zhang, M.”Full-image guided filtering for fast stereo matching”. In: *IEEE Signal Processing Letters*. 2013; 20 (3): 237–240. DOI: https://doi.org/10.1109/LSP.2013.2241759.

21. Yang, Q. “Local smoothness enforced cost volume regularization for fast stereo correspondence”. In: *IEEE Signal Processing Letters*. 2015; 22 (9): 1429–1433. DOI: https://doi.org/10.1109/LSP.2015.2409203.

22. Shi, J. & Liu, P. “A memory and computation efficient local stereo algorithm suited for mobile robot applications”. *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. Macau, China. 2020. p. 1–6. DOI: https://doi.org/10.1109/ICSPCC50002.2020.9259560.

23. Smirnov, S. & Gotchev, A. “Fast hierarchical cost volume aggregation for stereo-matching”. *IEEE Visual Communications and Image Processing Conference*. Valletta, Malta. 2014. p. 498–501. DOI: https://doi.org/10.1109/VCIP.2014.7051615.

24. Hirschmuller, H. “Stereo processing by semiglobal matching and mutual information”. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2007; 30 (2): 328–341. DOI: https://doi.org/10.1109/TPAMI.2007.1166.

25. Hosni, A., Rhemann, C., Bleyer, M., Rother, C. & Gelautz, M. “Fast cost-volume filtering for visual correspondence and beyond”. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2012; 35 (2): 504–511. DOI: https://doi.org/10.1109/TPAMI.2012.156.

26. Brox, T. & Malik, J. “Large displacement optical flow: descriptor matching in variational motion estimation”. *In IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010; 33 (3): 500–513. DOI: https://doi.org/10.1109/TPAMI.2010.143.

27. Kordelas, G. A., Alexiadis, D. S., Daras, P. & Izquierdo, E. “Content-based guided image filtering, weighted Semi-Global optimization, and efficient disparity refinement for fast and accurate disparity estimation”. *In IEEE Transactions on Multimedia*. 2015; 18 (2): 155–170. DOI: https://doi.org/10.1109/TMM.2015.2505905.

28. Liu, H., Liu, Y., OuYang, S., Liu, C. & Li, X.”A novel method for stereo matching using gabor feature image and confidence mask”. *Visual Communications and Image Processing (VCIP)*. Kuching, Malaysia. 2013. p. 1–6. DOI: https://doi.org/10.1109/VCIP.2013.6706388.

29. Mayer, N., *et al.* “A large dataset to train convolutional networks for disparity, optical flow, and scene flow estimation”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA. 2016. p. 4040–4048. DOI: https://doi.org/10.1109/CVPR.2016.438.

30. Girshick, R., Donahue, J., Darrell, T. & Malik, J. “Rich feature hierarchies for accurate object detection and semantic segmentation”. *IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA. 2014. p. 580–587. DOI: https://doi.org/10.1109/CVPR.2014.81.

31. Fidler, S., Mottaghi, R., Yuille, A. & Urtasun, R. “Bottom-Up segmentation for top-down detection”. *IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA. 2013. p. 3294–3301. DOI: https://doi.org/10.1109/CVPR.2013.423.

32. Dalal, N. & Triggs, B. “Histograms of oriented gradients for human detection”. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*. San Diego, CA, USA. 2005; 1: 886–893. DOI: https://doi.org/10.1109/CVPR.2005.177.

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

Financing: The study was conducted with financial support from grant No. 131/0161 of the National Research Foundation of Ukraine

Received 12.09.2024
Received after revision 15.11.2024
Accepted 21.11.2024

DOI: https://doi.org/10.15276/aait.07.2024.23
УДК 004.93

Методи уточнення карти глибин, отриманої з датчиків глибини

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@od.edu.ua. Scopus Author ID: 57188700773
[1)] Національний університет «Одеська Політехніка», пр. Шевченка, 1. Одеса, Україна, 65044

АНОТАЦІЯ

Карти глибини мають важливе значення для таких застосувань, як робототехніка, доповнена реальність, автономні транспортні засоби та медична візуалізація, надаючи критично важливу просторову інформацію. Однак карти глибини, отримані за допомогою таких датчиків, як датчики часу польоту (ToF) і системи структурованого світла, часто страждають від низької роздільної здатності, шуму і пропущених даних. Для вирішення цих проблем у цьому дослідженні представлено інноваційний метод уточнення карт глибини шляхом інтеграції кольорових зображень високої роздільної здатності. Запропонований підхід використовує як жорсткі, так і м'які стратегії розподілу пікселів для адаптивного покращення якості карти глибини. Модель з жорстким рішенням спрощує класифікацію країв, тоді як модель з м'яким рішенням, інтегрована в рамках теорії випадкових полів Маркова, покращує узгодженість країв і зменшує шум. Аналізуючи розбіжності між краями на картах глибини та кольорових зображеннях, метод ефективно усуває такі артефакти, як копіювання текстури та розмиті краї, забезпечуючи краще узгодження між наборами даних. Ключові інновації включають використання оператора виявлення країв Кенні для виявлення і класифікації невідповідностей країв та обчислення анізотропної спорідненості для точного структурного представлення. Модель з м'яким прийняттям рішень впроваджує передові методи зменшення шуму, покращуючи роздільну здатність карти глибини і зберігаючи деталі країв краще, ніж традиційні методи. Експериментальна перевірка на еталонних наборах даних Middlebury демонструє, що запропонований метод перевершує існуючі методи у зменшенні значень середньої абсолютної різниці, особливо у сценаріях з високим масштабуванням. Візуальне порівняння

підкреслює його здатність придушувати артефакти і підвищувати різкість країв, що підтверджує його ефективність у різних умовах. Цей підхід має значний потенціал для застосувань, що потребують високоякісних карт глибини, включаючи робототехніку, доповнену реальність, автономні системи та медичну візуалізацію. Усуваючи критичні обмеження існуючих методів, дослідження пропонує надійне, універсальне рішення для уточнення карт глибини з можливостями оптимізації в реальному часі в динамічних середовищах..

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. Національний університет «Одеська Політехніка», 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 комп'ютерних систем, кафедри Штучного інтелекту та аналізу даних. Національний університет «Одеська Політехніка», пр. Шевченка, 1. Одеса, 65044,Україна