Vol. 11, No. 1, Spring and Summer 2025, pp. 152-174 ISSN (Print): 2538-449X ISSN (Online): 2981-1619

Research Article
OPEN ACCESS
Kharazmi Journal of Earth Sciences
Journal homepage https://gnf.khu.ac.ir

Depth estimation and 3D reconstruction from a single image based on the MiDaS deep learning model

Mahdi Farhangi [1], Asghar Milan [2]*, Gholamreza Fallahi[2], Ehsan Khankeshi-Zadeh[3]

1, 2. Faculty of Civil Environmental and Water Engineering, Shahid Beheshti University, Tehran, Iran.
3. Faculty of Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran.

Article info

Article history
Received: 25 March 2025
Accepted: 16 August 2025

Keywords:
Depth Estimation, 3D Reconstruction, Single Image, Deep Learning, Photogrammetry, Machine Vision.

Abstract

3D reconstruction plays an important role in surveying and close-range photogrammetry, facilitating the accurate extraction of geometric information from objects and their surrounding environment. However, conventional methods in this field typically require multi-view images along with positional and angular data of the camera, which can pose limitations in certain practical applications. This study introduces a novel approach based on the MiDaS deep learning model, one of the most accurate architectures for monocular depth estimation, which is capable of generating a relative depth map from a single 2D image. The final 3D model is then extracted using the Poisson Surface Reconstruction algorithm, without the need for spatial information or camera orientation data. To evaluate the performance of the proposed method, the resulting 3D model was compared against a reference model produced by the conventional photogrammetry method. The results showed a Root Mean Square Error (RMSE) of 0.775 centimeters, confirming the appropriate accuracy and reliability of the proposed approach under multi-view data limitations. This study demonstrates the high potential of deep learning models like MiDaS in 3D reconstruction and surveying applications, and highlights that using more advanced versions such as DPT could further improve accuracy in future research and applications.

Introduction

Three-dimensional (3D) reconstruction and depth estimation from two-dimensional (2D) images are fundamental topics in fields such as surveying, photogrammetry, computer vision, and augmented reality (Transfer et al., 2020). These processes aim to extract precise geometric information from objects and environments, enabling applications in modeling, structural analysis, and design. Traditional 3D reconstruction methods, such as stereo vision and Structure from Motion (SfM), rely heavily on multiple images captured from different angles and precise positional data from imaging stations. This dependency poses significant operational challenges, particularly in scenarios where acquiring multi-view images is impractical or impossible, requiring complex equipment, time-consuming processes, and sensitivity to varying lighting conditions. In contrast, recent advances in deep learning have introduced innovative solutions for depth estimation and 3D reconstruction from a single image, mitigating these limitations. This study aims to develop and evaluate a novel deep learning-based approach for 3D reconstruction from a single 2D image. The proposed method utilizes the MiDaS (Multiple Depth Estimation Accuracy with Single Network) model for depth estimation, followed by Poisson surface reconstruction to generate the final 3D model (Wang et al., 2004). The primary objective is to assess the efficacy of this approach compared to multi-image photogrammetry, offering an efficient solution for constrained conditions.

Materials and Methods

The study utilized a dataset from the PhotoModeler software training collection, consisting of images captured with a FinePix F10 camera wit h a focal length of 8.1755 mm. For the photogrammetry method, four

DOI http://doi.org/10.22034/KJES.2025.11.1.107622
***Corresponding author:** Asghar Milan; **E-mail:** a_milan@sbu.ac.ir
How to cite this article: Farhangi , M., Milan , A., Fallahi, G. R., Khankeshi-Zadeh, E., 2025. Depth estimation and 3D reconstruction from a single image based on the MiDaS deep learning model. Kharazmi Journal of Earth Sciences 11(1), 152- 174. http://doi.org/10.22034/KJES.2025.11.1.107622

complementary images of the target object were used to create a high-precision reference 3D model. In contrast, the proposed deep learning method employed only a single frontal-view image. This image was captured at an oblique angle under optimized conditions, including uniform lighting (using LED panels), appropriate depth of field, and an optimal distance from the object to ensure sufficient geometric information. The MiDaS model, based on convolutional neural networks (CNNs) and trained on diverse datasets such as KITTI, NYU Depth V2, and Make3D, was applied to generate a depth map. The model configuration included the MiDaS_small version with four main layers (featuring components like Conv2dSameExport and InvertedResidual) and image preprocessing steps (resizing to 384 pixels, normalization, and tensor conversion).

The depth map produced by MiDaS was converted into a 3D point cloud using the pinhole camera model, where 2D pixel coordinates (x, y) and their corresponding depth values (v) were transformed into 3D coordinates (X, Y, Z) based on geometric equations involving the focal length and principal point. Surface normals were then computed using the K-D Tree algorithm to determine surface orientations, enhancing reconstruction accuracy. These normals, along with the point cloud, were fed into the Poisson surface reconstruction method, which solved the Poisson differential equation to produce a continuous and smooth 3D mesh. For comparison, the reference 3D model was generated using PhotoModeler from multi-view images. Both models (MiDaS and photogrammetry) were loaded into CloudCompare, where point-to-point distance measurements were conducted. The root mean square error (RMSE) was used to evaluate accuracy. All depth map processing and point cloud generation steps were implemented in Python within the Google Colab environment.

Results and Discussion

The results demonstrated that MiDaS successfully generated a relative depth map from the single 2D image, with brighter colors indicating closer objects and darker shades representing farther ones. This depth map was effectively transformed into a 3D point cloud, accurately reflecting the object's overall geometry. The Poisson surface reconstruction produced a smooth and continuous 3D mesh that preserved the object's details and boundaries. However, areas with low texture or edges exhibited reduced detail, influenced by the initial depth map quality and point cloud density. Comparison with the photogrammetry model revealed an RMSE of 0.775 cm (approximately 8 mm) for the MiDaS model, compared to 0.037 cm for the photogrammetry model. This minor difference underscores the acceptable accuracy of the proposed method, especially considering its reliance on a single image.

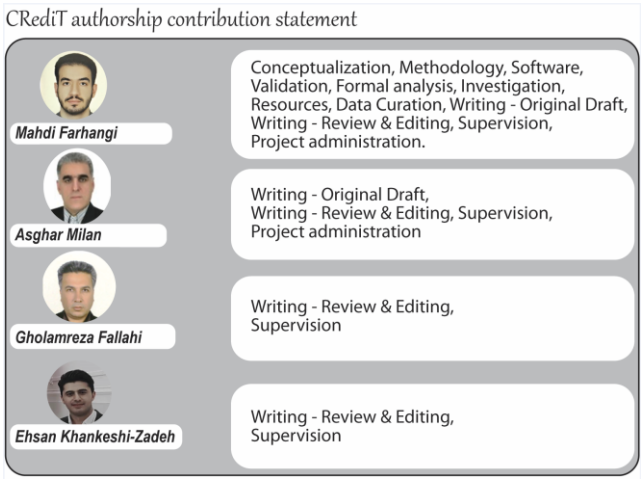
Discussion of the findings suggests that MiDaS's depth estimation performance depends on factors such as input image quality, training data diversity, and preprocessing settings. In regions with complex geometry or suboptimal lighting, depth accuracy may decline, impacting the 3D model. Conversely, photogrammetry benefits from multi-view data, enabling precise reconstruction of hidden areas and varied angles, albeit with greater complexity and cost. The key advantage of MiDaS lies in its simplicity, rapid processing, and independence from additional data, making it suitable for real-time applications or resource-limited scenarios. Nonetheless, the observed discrepancies indicate that for highly precise applications, such as industrial analysis, model optimization or advanced MiDaS versions may be necessary. Enhancing the Poisson reconstruction algorithm or training MiDaS with more diverse datasets could further narrow this gap, improving overall accuracy and detail retention.

Conclusions

This study confirmed that the MiDaS model, using a single image without camera positional data, can produce 3D models with acceptable accuracy (RMSE of 0.775 cm). Compared to the photogrammetry model (RMSE of 0.037 cm), the proposed method demonstrates efficacy in constrained conditions. The Poisson surface

[DOI: 10.22034/KJES.2025.11.1.107622]

reconstruction effectively generated smooth and precise meshes from the point cloud, though minor differences persist due to the inherent limitations of single-image depth estimation versus multi-view photogrammetry. The research highlights MiDaS as a viable option for applications prioritizing speed and simplicity, yet for scenarios demanding high precision, traditional methods or enhanced MiDaS versions are preferable. Future research should explore advanced MiDaS variants, hybrid single- and multi-view approaches, and improved reconstruction algorithms to boost accuracy and detail. This study provides a foundation for advancing deep learning applications in photogrammetry and computer vision, showcasing its potential in modern 3D reconstruction workflows.



References

Howells, S., Abuomar, O. 2022. Depth Maps Comparisons from Monocular Images by MiDaS Convolutional Neural Networks and Dense Prediction Transformers. In International Conference on Electrical, Computer, Communications and Mechatronics Engineering, ICECCME 2022 (pp. 1–6). IEEE. Retrieved from https://doi.org/10.1109/ICECCME55909.2022.9987767

Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612.

[DOI: 10.22034/KJES.2025.11.1.107622]

جلد ۱۱، شماره ۱، بهار و تابستان ۱۴۰۴، صفحه ۱۵۲ تا ۱۷۴ شاپا چاپی: ۴۴۹x- ۲۵۳۸ شاپا الکترونیکی: ۱۶۱۹-۲۹۸۱

مقاله پژوهشی

دسترسی آزاد

مجله علوم زمین خوارزمی

Journal homepage https://gnf.khu.ac.ir

تخمین عمق و بازسازی سه‌بعدی از تک تصویر مبتنی بر مدل یادگیری عمیق MiDaS

مهدی فرهنگی[۱]، اصغر میلان*[۲]، غلامرضا فلاحی[۲]، احسان خانکشی‌زاده[۳]

۱، ۲. دانشکده مهندسی عمران، آب و محیط زیست، دانشگاه شهید بهشتی، تهران، ایران.

۳. دانشکده مهندسی نقشه‌برداری، دانشگاه خواجه نصیرالدین طوسی، تهران، ایران.

اطلاعات مقاله

تاریخچه مقاله

دریافت: ۱۴۰۴/۰۱/۰۵

پذیرش: ۱۴۰۴/۰۵/۲۵

واژه‌های کلیدی

تخمین عمق، بازسازی سه‌بعدی، تک تصویر، یادگیری عمیق، فتوگرامتری، بینایی ماشین.

چکیده

بازسازی سه‌بعدی نقش مهمی در نقشه‌برداری و فتوگرامتری برد کوتاه ایفا می‌کند و به استخراج دقیق اطلاعات هندسی از اشیا و محیط اطراف کمک می‌نماید. روش‌های مرسوم این حوزه معمولاً به تصاویر چندنمایی و داده‌های موقعیت و زاویه تصویربرداری نیاز دارند که در برخی کاربردهای عملی با محدودیت‌هایی همراه است. در این پژوهش، روشی نوین مبتنی بر مدل یادگیری عمیق MiDaS، یکی از معماری‌های دقیق تخمین عمق تک‌نما، معرفی شده است که قادر است تنها با یک تصویر دوبعدی، نقشه عمق نسبی تولید کند. سپس با بهره‌گیری از الگوریتم بازسازی سطح پواسون (Poisson Surface Reconstruction)، مدل سه‌بعدی نهایی بدون نیاز به اطلاعات مکانی یا زاویه دوربین استخراج می‌شود. برای ارزیابی عملکرد روش پیشنهادی، مدل سه‌بعدی حاصل با مدل مرجع تولید شده توسط روش رایج فتوگرامتری مقایسه شد. نتایج نشان داد که خطای میانگین مربعات (RMSE) برابر با ۰٫۷۷۵ سانتی‌متر است که دقت مناسب و قابلیت اعتماد روش پیشنهادی را در شرایط محدودیت داده‌های چندنمایی تأیید می‌کند. این مطالعه ظرفیت بالای مدل‌های یادگیری عمیق مانند MiDaS را در بازسازی سه‌بعدی و کاربردهای نقشه‌برداری و فتوگرامتری نشان می‌دهد و اشاره می‌کند که استفاده از نسخه‌های پیشرفته‌تر همچون DPT می‌تواند دقت نتایج را در پژوهش‌ها و کاربردهای آینده بهبود بخشد.

مقدمه

بازسازی سه‌بعدی و تخمین عمق از تصاویر دوبعدی، به‌عنوان یکی از موضوعات چالش‌برانگیز و مهم در زمینه‌های مختلف از جمله بینایی ماشین، فتوگرامتری، نقشه‌برداری و واقعیت افزوده (Augmented Reality) به شمار می‌رود. تک تصویر، به دلیل نداشتن اطلاعات عمق، به خودی خود قادر به ارائه داده‌های کافی برای مدل‌سازی سه‌بعدی دقیق نیستند. این مشکل، پژوهشگران را بر آن داشته تا از تکنیک‌های پیشرفته مانند یادگیری عمیق برای استخراج اطلاعات عمق از یک تصویر استفاده کنند. با ظهور مدل‌های مبتنی بر شبکه‌های عصبی عمیق، روش‌های جدیدی توسعه یافته‌اند که امکان تخمین عمق و بازسازی سه‌بعدی را تنها با استفاده از یک تصویر دوبعدی فراهم می‌آورند. در روش‌های سنتی بازسازی سه‌بعدی، مانند استریو ویژن (Stereo Vision)، شامل استفاده از دو یا چند دوربین برای تحلیل اختلاف دید بین تصاویر ثبت شده است (Howells and Abuomar, 2022). همچنین روش ساختار از حرکت (Structure from Motion (SfM))، به فرآیند بازسازی ساختار سه‌بعدی از یک مجموعه تصاویر

DOI http://doi.org/10.22034/KJES.2025.11.1.107622

*نویسنده مسئول: اصغر میلان a_milan@sbu.ac.ir

استناد به این مقاله: فرهنگی، م.، میلان، ا.، فلاحی، غ. ر.، خانکشی‌زاده، ا. (۱۴۰۴). تخمین عمق و بازسازی سه‌بعدی از تک تصویر مبتنی بر مدل یادگیری عمیق MiDaS. مجله علوم زمین خوارزمی. جلد ۱۱، شماره ۱، صفحه ۱۵۲ تا ۱۷۴. http://doi.org/10.22034/KJES.2025.11.1.107622

اشاره دارد، نیازمند چندین نما از یک صحنه یا شیء هستند تا عمق و ساختار سه‌بعدی تخمین زده شود(Fusiello, 2024). هرچند این روش‌ها در برخی کاربردها به موفقیت‌هایی دست یافته‌اند، اما محدودیت‌های بسیاری مانند نیاز به چندین تصویر دقیق، پیچیدگی در تنظیم ایستگاه‌های تصویربرداری و پردازش‌های محاسباتی سنگین وجود دارد. علاوه بر این، روش‌های چند تصویری به دلیل نیاز به تصاویر متعدد، در محیط‌های پیچیده یا شرایط نوری متغیر با مشکلاتی مواجه می‌شوند که به کاهش دقت و کارایی آن‌ها منجر می‌گردد.

در راستای غلبه بر این چالش‌ها، بازسازی سه‌بعدی و تخمین عمق از روی تک تصویر به‌عنوان یک راهکار نوین مطرح شده‌اند. یکی از این راهکارها استفاده از الگوریتم‌های یادگیری عمیق است. شبکه‌های عصبی عمیق با قابلیت استخراج ویژگی‌های پیچیده از تصاویر دوبعدی، پتانسیل بالایی در تخمین عمق و بازسازی سه‌بعدی دقیق ارائه می‌دهند(Piccinelli et al., 2024). یکی از مدل‌های برجسته در این زمینه MiDaS (Multiple Depth Estimation Accuracy with Single Network) است. الگوریتم MiDaS با بهره‌گیری از شبکه‌های عصبی عمیق، امکان تخمین عمق دقیق از تک تصویر را فراهم می‌کند. این الگوریتم با آموزش روی مجموعه داده‌های گسترده‌ای مانند KITTI، NYU Depth و Make3D، توانسته است عملکرد قابل‌توجهی در شرایط مختلف ارائه دهد. این مدل به‌ویژه در شرایطی که داده‌های چند تصویری موجود نیستند یا استفاده از آن‌ها امکان‌پذیر نیست، به عنوان یک راهکار کارآمد مورد توجه قرار گرفته است(Ranftl et al., 2020). پیشرفت‌های اخیر در بازسازی سه‌بعدی و تخمین عمق از تک تصویر، نوآوری‌های قابل توجهی را در این دو حوزه به همراه داشته است. یکی از اولین تلاش‌ها در این زمینه توسط ساکسنا و همکاران (Saxena et al., 2007) در سال‌های ۲۰۰۷ و ۲۰۰۸ انجام شد. آن‌ها رویکردی نظارت شده برای تخمین عمق از یک تصویر ارائه دادند که از میدان تصادفی مارکوف (Markov Random Field (MRF) چند مقیاسه استفاده می‌کند و قادر به بازسازی دقیق نقشه‌های عمق در صحنه‌های غیرساختار یافته می‌باشد. در تحقیق دیگری از این نویسنده، تخمین عمق سه‌بعدی از تصاویر تک‌چشمی با استفاده از یک روش یادگیری ماشین نظارت‌شده انجام شد. مدل پیشنهادی مبتنی بر میدان تصادفی مارکوف چند مقیاسی است که ویژگی‌های محلی و سراسری تصویر را ترکیب کرده و قادر به تولید نقشه‌های عمق دقیق برای صحنه‌های غیرساختار یافته است. همچنین طبق گزارش آن‌ها، با ترکیب تصاویر تک و استریو، دقت تخمین عمق بهبود یافت(Saxena et al., 2008). ایگن و همکاران (Eigen et al., 2014) روشی جدید برای پیش‌بینی عمق از تک تصویر ارائه دادند که از دو شبکه عمیق برای پیش‌بینی سراسری و تصحیح محلی استفاده می‌کند. آن‌ها از مجموعه داده‌های NYU Depth (New York University Depth Dataset) که شامل صحنه‌هایی از فضای داخلی با وضوح بالا همراه با نقشه‌های عمق واقعی تهیه‌شده توسط دانشگاه نیویورک است (Silberman et al., 2012) و همچنین از مجموعه داده‌های KITTI (Karlsruhe Institute of Technology and Toyota Technological Institute) که توسط موسسه فناوری کارلسرا و موسسه (Max Planck Institute(MPI)) برای سیستم‌های هوشمند در آلمان تهیه شده و شامل تصاویر فضای باز و داده‌های عمق مرتبط با سیستم‌های رانندگی خودکار است (Eldesokey et al., 2020)، برای آموزش و ارزیابی مدل خود بهره گرفتند. این روش در هر دو مجموعه داده نتایج برجسته‌ای به دست آورد و قادر به تعیین دقیق مرزهای عمق بدون نیاز به قطعه‌بندی تصویر گردید. هرمن و همکاران (Hermann et al., 2020)، روشی خود-نظارتی (Self-Supervised Method) برای تخمین عمق از تصاویر هوایی معرفی کردند که نیازی به داده‌های آموزشی با برچسب ندارد و دقت بالایی را در تخمین عمق و موقعیت ارائه می‌دهد. خان و همکاران (Khan et al., 2020) مروری بر روش‌های مبتنی بر یادگیری عمیق برای تخمین عمق از تصاویر ((Red-Green-Blue(RGB) ارائه دادند و به بررسی داده‌های مرتبط و ۱۳ روش پیشرفته در این حوزه پرداختند و چالش‌های آینده را مورد بحث قرار دادند. در جدول (۱)، روش‌های مذکور در این مطالعه آمده است.

[DOI: 10.22034/KJES.2025.11.1.107622]

جدول ۱- روش‌های تخمین عمق مبتنی بر مدل‌های یادگیری عمیق

Table 1 - Depth Estimation Methods Based on Deep Learning Models

روش	دسته‌بندی	معماری	مجموعه داده	بهینه‌ساز	چالش‌ها	حوزه کاربرد
EMDEOM	نظارت شده	FC	KITTI	Adam	حساسیت به نویز، محدودیت داده‌های واقعی	خودروهای خودران
ACAN	نظارت شده	Encoder-Decoder	NYU-v2, KITTI	SGD	زمان آموزش بالا، نیاز به داده زیاد	بینایی ماشین، رباتیک
DenseDepth	نظارت شده	Encoder-Decoder	NYU-v2	Adam	عدم تعمیم خوب به محیط‌های غیرآموزشی	رباتیک، واقعیت افزوده
DORN	نظارت شده	CNN	NYU-v2, KITTI	Adam	پیچیدگی محاسباتی بالا	سیستم‌های کمک راننده
VNL	نظارت شده	Encoder-Decoder	NYU-v2, KITTI	SGD	وابستگی به داده‌های دقیق عمق	بینایی کامپیوتر
BTS	نظارت شده	Encoder-Decoder	NYU-v2, KITTI	Adam	حساسیت به بازنمایی لبه‌ها	بازسازی سه بعدی، واقعیت مجازی
DeepV2D	نظارت شده	CNN	NYU-v2, KITTI	Adam	نیاز به تصاویر چندگانه، هزینه محاسباتی	بازسازی سه بعدی دینامیک
LISM	خود نظارتی	Encoder-Decoder	KITTI	Adam	دقت کمتر نسبت به روش‌های نظارت‌شده	کاربردهای واقعیت افزوده
monoResMatch	خود نظارتی	CNN	KITTI	Adam	محدودیت در شرایط نوری ضعیف	نقشه‌برداری موبایل
PackNet-SfM	خود نظارتی	CNN	KITTI	Adam	وابستگی به صحت تخمین حرکت دوربین	رباتیک
monodepth2	خود نظارتی	CNN	KITTI	Adam	چالش‌های تعمیم‌پذیری	خودروهای خودران
GASDA	نیمه نظارتی	CNN	KITTI	Adam	تعادل بین داده نظارت‌شده و بدون نظارت	بینایی ماشین
VOMonodepth	خود نظارتی	Auto-Decoder	KITTI	Adam	پایداری در صحنه‌های پویا	کاربردهای تخمین عمق پویا

در تحقیقات جدیدتر، استفاده از شبکه‌های عصبی عمیق برای تخمین عمق گسترش یافته است. برای مثال، مینگ و همکاران (Ming et al., 2021) به مرور توسعه‌های اخیر در تخمین عمق با استفاده از یادگیری عمیق پرداختند و به بررسی مدل‌های یادگیری عمیق، دسته‌بندی روش‌ها و چالش‌های موجود پرداختند. هریستوا و همکاران (Hristova et al., 2022) مدلی برای پیش‌بینی عمق از یک تصویر در محیط‌های جنگلی ارائه دادند. مدل آن‌ها با استفاده از مجموعه داده جدید از تصاویر Red-Green-Blue-Depth(RGB-D)، دقت پیش‌بینی را بهبود بخشید و خطای میانگین مربعات (Root Mean Square Error (RMSE)) مدل را به ۲/۱ متر کاهش داد. ولپونر و همکاران (Welponer et al., 2022) به ارزیابی معماری‌های شبکه عصبی پیچشی نظارت شده برای تخمین عمق از تک تصویر پرداختند. آن‌ها به محدودیت‌های این روش‌ها در بازسازی ساختار سه‌بعدی دقیق اشاره کرده و یک معیار جدید به نام ArchDepth را برای ارزیابی عملکرد این روش‌ها معرفی کردند.

[DOI: 10.22034/KJES.2025.11.1.107622]

نجف و همکاران (Najaf et al., 2023) به بررسی تخمین عمق از تصاویر ماهواره‌ای Google Earth با استفاده از شبکه‌های عصبی پیچشی پرداختند. آن‌ها یک شبکه عصبی کانولوشن را برای تولید نقشه عمق با وضوح بالا از تصاویر RGB معرفی کردند و نتایج قابل قبولی ارائه دادند که با تصاویر معیار (International Society for Photogrammetry and Remote Sensing (ISPRS)) قابل مقایسه است. راجاپاکشا و همکاران (Rajapaksha et al., 2024)، مقاله‌ای جامع درباره روش‌های تخمین عمق بر اساس یادگیری عمیق از تک تصویر و ویدیوها ارائه دادند و به بررسی پیشرفت‌ها، دسته‌بندی روش‌ها و چالش‌های موجود پرداختند.

در زمینه بازسازی سه‌بعدی نیز، کارهای متعددی انجام شده است. ازدن و همکاران (Ozden et al., 2007) به چالش‌های بازسازی اطلاعات سه‌بعدی صحنه و شناسایی مسیر حرکت دوربین از توالی تصاویر پرداخته‌اند. آن‌ها الگوریتمی ارائه کرده‌اند که می‌تواند تغییرات در صحنه، مانند ورود یا خروج اشیاء، جدا شدن مسیر حرکتی اشیاء مشترک، یا ادغام مسیرهای حرکتی اشیاء مختلف را شناسایی و مدیریت کند. این الگوریتم به‌طور هم‌زمان ویژگی‌های تصویر را دنبال کرده، آن‌ها را به گروه‌های متحرک و صلب تقسیم‌بندی می‌کند و تمامی بخش‌ها را به‌صورت سه‌بعدی بازسازی می‌کند. روش پیشنهادی، برخلاف تکنیک‌های پردازش دسته‌ای، به‌صورت برخط (Online) عمل کرده و برای بازسازی اشیاء کوچک و متحرک در پیش‌زمینه بسیار کارآمد است. سالزمان و فوآ (Salzmann and Fua, 2010) در کتاب خود روش‌هایی را بررسی کرده‌اند که امکان بازسازی شکل سطوح سه‌بعدی تغییرپذیر را از یک جریان ویدئویی فراهم می‌کنند. آن‌ها دو دسته روش مؤثر را معرفی کرده‌اند:

روش‌های مبتنی بر الگو که با ایجاد تطابق با یک تصویر مرجع از پیش شناخته‌شده عمل می‌کنند.

روش‌های ساختار از حرکت غیر صلب (Non-Rigid Structure-from-Motion (NRSfM)) که با دنبال کردن نقاط در طول ویدئو، شکلی ناشناخته را بازسازی می‌کنند.

آن‌ها همچنین ابهامات ذاتی این روش‌ها را بررسی و راه‌حل‌های عملی برای رفع این ابهامات ارائه کرده‌اند. این مطالعه، مسیرهایی را نیز برای تحقیقات آینده پیشنهاد می‌دهد. لانسچر و همکاران (Lunscher and Zelek, 2018) در پژوهش خود به بررسی کاربردهای اسکن سه‌بعدی بدن پرداختند. آن‌ها نشان دادند که با استفاده از یادگیری عمیق، می‌توان اسکنرهای ساده‌ای طراحی کرد که تنها به یک نقشه عمق ورودی برای تولید ابر نقاط تمام بدن نیاز دارند. این اسکنرها می‌توانند در زمینه‌های مختلفی مانند مد، پوشاک و نظارت بر سلامت استفاده شوند. مدل آن‌ها قادر به تولید ابر نقاط با دقت ۵/۱۹ میلی‌متر است و می‌تواند به کاهش هزینه‌ها در استفاده از سیستم‌های اسکن بدن کمک کند. شو و همکاران (Chu et al., 2019) یک روش برای تولید ابر نقاط سه‌بعدی از یک تصویر RGB ارائه دادند که شامل تولید تصویر عمق با استفاده از شبکه مولد تخاصمی (Generative Adversarial Network (GAN)) و سپس محاسبه ابر نقاط از تصویر عمق است. این روش نشان‌دهنده تولید ابر نقاط با کیفیت بالا و نیاز به منابع محاسباتی کم است. شبکه مولد ضدی یک نوع معماری شبکه عصبی در یادگیری عمیق است که به‌طور خاص برای تولید داده‌های جدید، مانند تصاویر، طراحی شده است. چو و همکاران (Choe et al., 2021) در مقاله خود Volume Fusion را معرفی کردند، روشی برای ادغام تخمین عمق و یادگیری عمیق در بازسازی صحنه‌های سه‌بعدی که عملکرد بهتری نسبت به تکنیک‌های سنتی و یادگیری عمیق موجود نشان می‌دهد.

در مطالعه‌ای، مدهاناند و همکاران (Madhuanand et al., 2021) از یادگیری خود-نظارتی برای تخمین عمق از ویدیوهای

[DOI: 10.22034/KJES.2025.11.1.107622]

UAVاستفاده کردند و مدل خود را با استفاده از فریم‌های ویدیویی و شبکه‌های عصبی پیچشی دو و سه بعدی آموزش دادند. نتایج آن‌ها نشان‌دهنده عملکرد بهتری نسبت به روش‌های موجود در تخمین عمق است. وندت و همکاران (Wandt et al., 2016) به بررسی تخمین شکل و حرکت سه‌بعدی انسانی از توالی تک تصویر پرداختند و مدلی را معرفی کردند که با استفاده از مقادیر پایه و فرضیات دوره‌ای توانسته است بازسازی سه‌بعدی دقیقی ارائه دهد.

تخمین عمق و بازسازی سه‌بعدی دو حوزه کاملاً مجزا هستند و هر یک از این حوزه‌ها شامل روش‌ها و تکنیک‌های خاص خود هستند. همانطور که گفته شد، برای تخمین عمق، از روش‌هایی مانند استریو ویژن، ساختار از حرکت (SfM) و شبکه‌های عصبی عمیق استفاده می‌شود، در حالی که برای بازسازی سه‌بعدی روش‌هایی نظیر بازسازی سطح پواسون (Poisson Surface Reconstruction)، مثلث‌بندی دلونی (Delaunay Triangulation) و روش‌های چند تصویری مورد استفاده قرار می‌گیرند(Cazals and Giesen, 2004). وجه تمایز و نوآوری این تحقیق در استفاده از یک مدل یادگیری عمیق تخمین عمق (MiDaS) برای بازسازی سه‌بعدی دقیق از تنها یک تصویر دوبعدی است، بدون نیاز به اطلاعات موقعیت دوربین یا داده‌های چند نمایی که در روش‌های سنتی فتوگرامتری ضروری هستند. برخلاف بسیاری از پژوهش‌های پیشین که تنها به استخراج نقشه عمق و کاربرد در حوزه بینایی ماشین بسنده کرده‌اند، در این تحقیق با دیدگاه کاربرد در مهندسی نقشه‌برداری، از نقشه عمق حاصل برای تولید مدل سه‌بعدی و انجام ارزیابی کمّی با استفاده از معیار RMSE نسبت به مدل دقیق حاصل از فتوگرامتری چند تصویری استفاده شده است. این روش به‌طور محسوسی نشان می‌دهد که می‌توان با اتکا به یک تصویر و بدون نیاز به سخت‌افزار خاص، بازسازی سه‌بعدی نسبتاً دقیقی انجام داد. این موضوع می‌تواند کاربردهای مهمی در محیط‌های دارای محدودیت منابع یا داده داشته باشد. نتایج این مطالعه می‌تواند به شناسایی مناسب‌ترین روش در کاربردهایی که با محدودیت داده‌های چند نمایی روبرو هستند، کمک کند. این رویکرد می‌تواند به توسعه کاربردهای متنوعی در فتوگرامتری، بینایی ماشین و سایر حوزه‌های مرتبط منجر شود و به عنوان یک مبنای جدید برای تحقیقات آینده در این زمینه‌ها مطرح گردد.

مواد و روش‌ها

در این تحقیق با هدف بازسازی سه‌بعدی و مقایسه روش مبتنی بر فتوگرامتری(به کارگیری چند تصویر) با مدل یادگیری عمیق MiDaS، از مجموعه داده آموزشی نرم‌افزار PhotoModeler بهره گرفته شده است. داده‌های تصویری با دوربین FinePix F10 تهیه شدند. این دوربین با فاصله کانونی دقیق ۸/۱۷۵۵ میلی‌متر که پس از کالیبراسیون دوربین به‌دست آمده، امکان تصویربرداری با وضوح بالا را برای بازسازی سه‌بعدی فراهم می‌آورد.

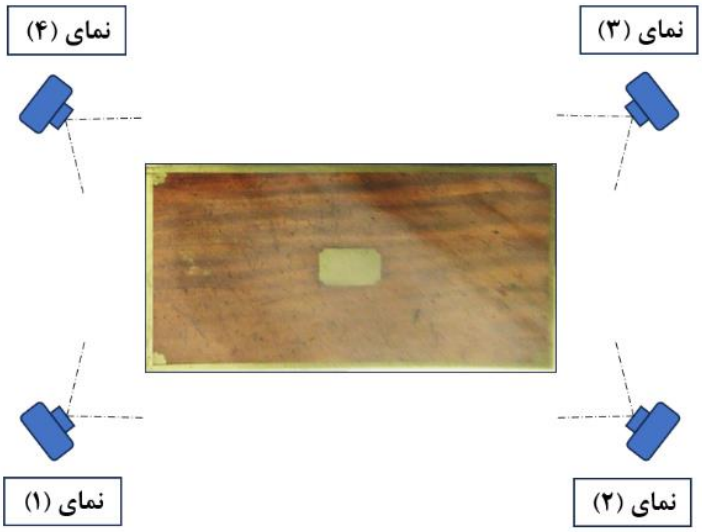
برای تهیه این مجموعه داده، فرآیند تصویربرداری به‌گونه‌ای طراحی شده است ثبت تصاویر از چهار زاویه مکمل نسبت به شیء صورت گیرد. این امر به بازسازی دقیق‌تر ابعاد و زوایای مختلف کمک می‌کند. تصاویر موجود در مجموعه داده، در شکل (۱) نشان داده شده است. این تصاویر با توجه به استانداردهای دقیق تصویربرداری و تنظیمات مناسب دوربین، به‌گونه‌ای ثبت شده‌اند که دقت لازم برای بازسازی را فراهم کنند.

[DOI: 10.22034/KJES.2025.11.1.107622]

شکل ۱- تصاویر اخذ شده با دوربین FinePix F10

Fig. 1. Images captured with the FinePix F10 camera

همانگونه که اشاره گردید انتخاب زوایای مناسب به بهبود کیفیت بازسازی و استخراج ویژگی‌های موردنظر کمک می‌کند. این زوایا در شکل (۲) نمایش داده شده‌اند.



شکل ۲- زوایای ثبت تصاویر

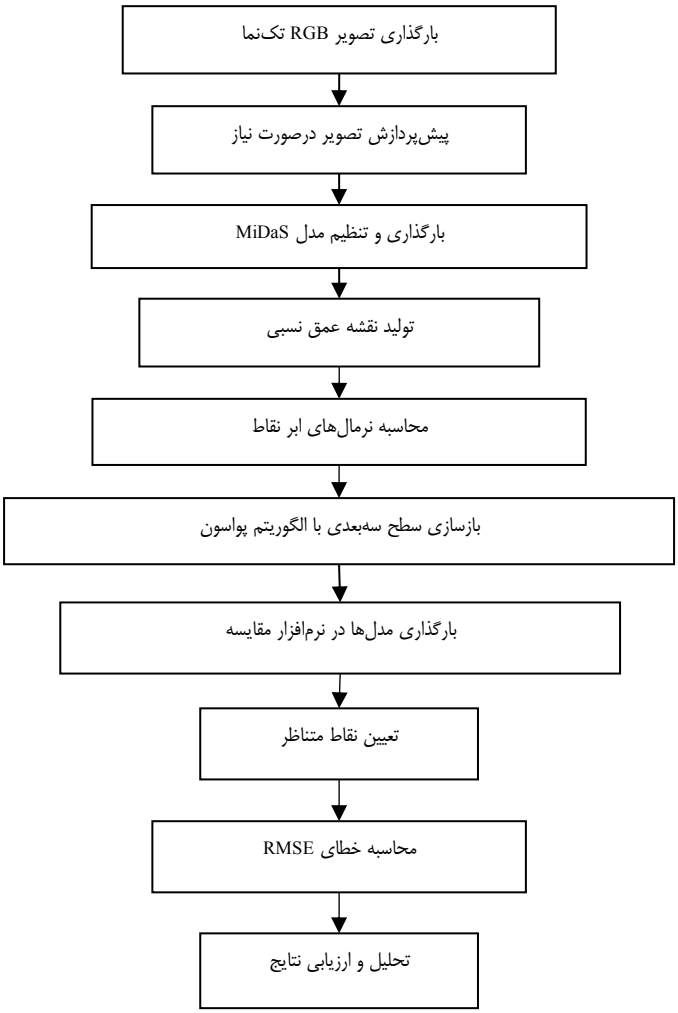
Fig. 2. Angles of image capture

طبق توضیحات مجموعه داده‌های آموزشی نرم‌افزار PhotoModeler، تنظیمات دوربین و شرایط محیطی به‌گونه‌ای انتخاب شده‌اند که داده‌های دریافتی دارای کیفیت و انسجام لازم برای مراحل پردازش و بازسازی سه‌بعدی باشند. در این تحقیق، برای بازسازی سه‌بعدی شیء به روش فتوگرامتری، از تمامی تصاویر موجود در مجموعه داده استفاده شد. همچنین، برای بازسازی سه‌بعدی با استفاده از مدل یادگیری عمیق MiDaS، تنها تصویر نمای اول به‌کار گرفته شد. این روند به ارزیابی کارایی MiDaS در بازسازی سه‌بعدی از تک تصویر کمک می‌کند.

در این پژوهش، روشی نوین برای بازسازی مدل‌های سه‌بعدی با استفاده از یادگیری عمیق ارائه شده است. فلوچارت موجود در شکل

[DOI: 10.22034/KJES.2025.11.1.107622]

(۳)، روند تحقیق و مراحل کلیدی پردازش داده‌ها را به‌صورت خلاصه نشان می‌دهد.



شکل ۳- مراحل انجام پژوهش

Fig. 3. Stages of the research process

تصویربرداری و شرایط آن در بازسازی سه‌بعدی

در این پژوهش، برای تهیه مدل سه‌بعدی با استفاده از مدل یادگیری عمیقMiDaS ، تنها یک تصویر دوبعدی از نمای جلوی شیء موردنظر انتخاب شد. این تصویر از مجموعه داده‌هایی استخراج شده است که به‌طور خاص برای بازسازی سه‌بعدی به روش فتوگرامتری طراحی و تهیه شده‌اند. با این حال، تصاویر این مجموعه داده با رعایت اصول و ملاحظات لازم، به‌گونه‌ای ثبت شده‌اند که امکان استفاده از آن‌ها در روش یادگیری عمیق نیز فراهم باشد. این ویژگی‌ها، دلیل انتخاب این مجموعه داده برای پژوهش حاضر بوده است.

✓ زاویه تصویربرداری

تصویر استفاده‌شده با زاویه‌ای مایل نسبت به شئ ثبت شده است. با توجه به اینکه تنها از یک تصویر برای تهیه مدل سه‌بعدی استفاده می‌شود، انتخاب زاویه تصویربرداری به‌گونه‌ای که بیشترین وجوه ممکن

[DOI: 10.22034/KJES.2025.11.1.107622]

شئ در تصویر ثبت گردد، اهمیت بالایی دارد. این اقدام اطلاعات بیشتری از هندسه شئ فراهم می‌کند و بازسازی سه‌بعدی دقیق‌تری ممکن می‌سازد. البته باید توجه داشت که نواحی پنهان شئ که توسط دوربین قابل مشاهده نیستند، قاعدتاً در مدل سه‌بعدی بازسازی نخواهند شد.

- ✓ **عمق میدان**

با توجه به شکل (۱)، برداشت می‌شود که برای ثبت جزئیات دقیق شئ، تنظیم عمق میدان (Depth of Field) با انتخاب دیافراگم مناسب انجام شده است. این تنظیم به‌گونه‌ای صورت گرفته که تمام بخش‌های تصویر، از نزدیک‌ترین تا دورترین نقاط، به‌صورت واضح ثبت شده‌اند. وضوح کامل تصویر ورودی از اهمیت بالایی برخوردار است، زیرا دقت نقشه عمق و در نهایت مدل سه‌بعدی به کیفیت تصویر وابسته است.

- ✓ **کیفیت نوری و کنترل انعکاس‌ها**

برای تصویربرداری دقیق و کاهش اثرات مزاحم نور، نورپردازی باید یکنواخت باشد تا از ایجاد سایه‌ها و انعکاس‌های ناخواسته جلوگیری شود. استفاده از منابع نور گسترده و نرم، مانند پنل‌های LED، به پخش یکنواخت نور و کاهش سایه‌های تیز کمک می‌کند. تنظیم مناسب شدت نور و حذف منابع نوری ناخواسته، مانند نور خورشید مستقیم، از اهمیت بالایی برخوردار است.

- ✓ **تنظیم فاصله دوربین تا شئ**

فاصله دوربین بر اساس فاصله کانونی دوربین (۸/۱۷۵۵ میلی‌متر) و اندازه شئ تعیین شده است تا جزئیات مهم شئ به‌خوبی در تصویر ثبت شوند. رعایت فاصله مناسب از شئ، تأثیر اعوجاج ناشی از لنز و تاری در تصویر را کاهش داده و کیفیت داده ورودی را تضمین می‌کند.

مدل یادگیری عمیق MiDaS و تنظیم آن

مدل MiDaS یکی از پیشرفته‌ترین مدل‌های یادگیری عمیق برای تخمین عمق از تک تصویر است که بدون نیاز به داده‌های چند نمایی می‌تواند نقشه عمق را با دقت بالا تولید کند. MiDaS بر مبنای معماری شبکه‌های عصبی پیچشی (Convolutional Neural Networks (CNN)) با مجموعه داده‌های KITTI، NYU Depth V2 و Make3D (Puscas et al., 2019)، آموزش و توسعه یافته و به‌ویژه برای پردازش تصاویر با شرایط نوری و محیط‌های پیچیده بهینه‌سازی شده است. این مدل، برخلاف روش‌های کلاسیک تخمین عمق که نیازمند چندین تصویر از زوایای مختلف یا اطلاعات موقعیت دوربین هستند، می‌تواند تنها با استفاده از یک تصویر دوبعدی، تصویر عمق دقیقی از صحنه ارائه دهد(Ranftl et al., 2020).

یکی از ویژگی‌های کلیدی MiDaS این است که با استفاده از داده‌های متنوع و گسترده‌ای آموزش دیده که شامل تصاویر از محیط‌های مختلف و شرایط متنوع است. این تنوع داده، مدل را قادر می‌سازد تا در شرایط مختلف و حتی با تغییرات پیچیده هندسی و نوری، عمق را به‌صورت نسبی تخمین بزند. این تخمین نسبی، اجسام نزدیک و دور را بر اساس میزان عمق از یکدیگر تفکیک کرده و به هر پیکسل یک مقدار عمق نسبی اختصاص می‌دهد. از دیگر مزایای MiDaS، بهره‌گیری از معماری چند مقیاسی است که امکان پردازش ویژگی‌های تصویر را در مقیاس‌های مختلف فراهم می‌کند. این ویژگی به مدل امکان می‌دهد که جزئیات بیشتری را در نقشه عمق تشخیص داده و لبه‌های اشیا را با دقت بیشتری مشخص کند، به‌ویژه در مناطقی که دارای جزئیات ریز یا تغییرات سریع در عمق هستند.

کاربرد MiDaS در این پژوهش، توانایی این مدل را در بازسازی سه‌بعدی صحنه‌های دوبعدی تنها از یک تصویر نشان می‌دهد و مزایای آن شامل ساده‌سازی روند بازسازی، کاهش نیاز به تنظیمات پیچیده تصویربرداری و کاهش هزینه‌های پردازشی است. در ادامه تنظیمات صورت گرفته بر روی مدل MiDaS، در جدول (۲) آمده است.

[DOI: 10.22034/KJES.2025.11.1.107622]

جدول ۲- تنظیمات مدل یادگیری عمیق MiDaS

Table 2 - Configuration settings of the MiDaS deep learning model

تنظیمات	جزئیات
مدل	MiDaS_small
تعداد لایه‌ها	۴ لایه اصلی
مؤلفه‌های اصلی	✓ Conv2dSameExport ✓ BatchNorm2d ✓ ReLU6 ✓ Depthwise Separable Conv ✓ Inverted Residual
پیش‌پردازش تصویر	✓ تغییر اندازه به ۳۸۴ پیکسل ✓ برش مرکز تصویر ✓ تبدیل به تنسور ✓ نرمال‌سازی با میانگین [۰,۴۵، ۰,۴۳، ۰,۴۰] و انحراف معیار [۰,۲۲، ۰,۲۲، ۰,۲۲]
بهینه‌ساز	Adam
خروجی	نقشه عمق

مدل MiDaS معمولاً شامل چهار لایه اصلی است که هرکدام دارای مؤلفه‌های مختلف هستند. این لایه‌ها به شرح زیر هستند:

لایه اول:

- Conv2dSameExport: این مؤلفه شامل لایه‌های کانولوشنی برای استخراج ویژگی‌های ابتدایی از تصویر ورودی است.
- BatchNorm2d و ReLU6: برای نرمال‌سازی و فعال‌سازی استفاده می‌شوند.
- Depthwise Separable Conv و Inverted Residual: برای کاهش پیچیدگی محاسباتی و استخراج ویژگی‌های عمیق‌تر.

لایه دوم:

- Inverted Residual: این مؤلفه شامل لایه‌های پیچیده‌تری هستند که به مدل کمک می‌کنند تا ویژگی‌های پیچیده‌تری را از تصویر استخراج کند.

لایه سوم:

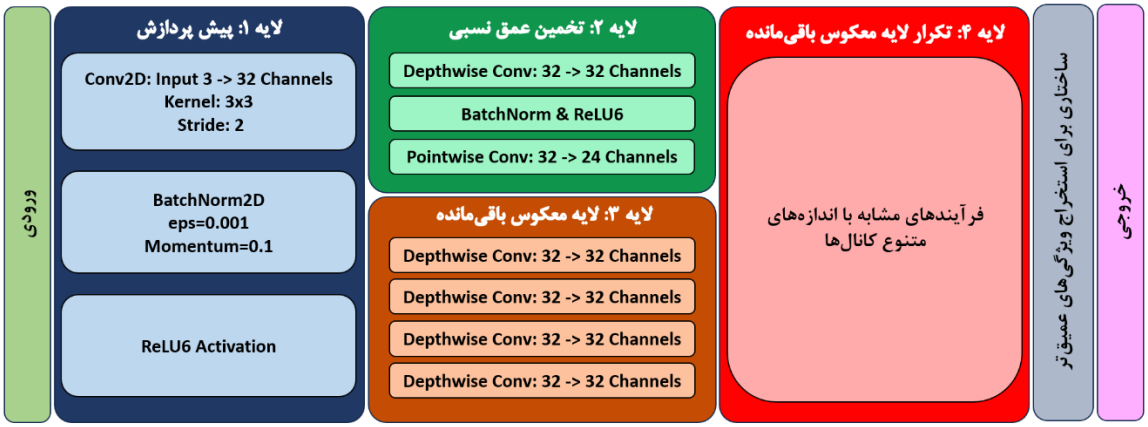
- Inverted Residual: ادامه‌ی لایه‌های پیچیده که با استفاده از لایه‌های کانولوشنی و نرمال‌سازی به مدل امکان استخراج ویژگی‌های پیچیده‌تر و پردازش بهتر اطلاعات را می‌دهد.

لایه چهارم:

- Inverted Residual: این لایه شامل لایه‌های پیشرفته‌تری برای استخراج ویژگی‌های عمقی‌تر و نهایی‌سازی نتایج عمق است.

شکل (۴)، نمای کلی معماری این مدل را نشان می‌دهد که ساختار لایه‌ها و مؤلفه‌ها را برای تخمین عمق دقیق و با کیفیت بالا به تصویر می‌کشد.

[DOI: 10.22034/KJES.2025.11.1.107622]



شکل ۴- معماری مدل یادگیری عمیق MiDaS

Fig. 4. Architecture of the MiDaS deep learning model

این ساختار لایه‌ها و مؤلفه‌ها به مدل MiDaS این امکان را می‌دهد که عمق را از تصاویر ورودی به‌طور دقیق و با کیفیت بالا تخمین بزند.

ایجاد نقشه عمق

در این مرحله، همان‌طور که اشاره شد، تصویر نمای اول شیء به عنوان ورودی مدل MiDaS استفاده می‌شود و نقشه عمق آن به عنوان خروجی مدل استخراج می‌شود. تصویر عمق تولیدشده توسط مدل MiDaS مقادیر عمق مطلق ارائه نمی‌دهد، بلکه نسبت‌های عمقی بین اجسام مختلف در تصویر را نشان می‌دهد، به‌طوری که اجسام نزدیک‌تر دارای عمق کمتر و اجسام دورتر دارای عمق بیشتری هستند. مقیاس عمق در این تصویر ممکن است ناهمگن باشد و برای نقاط مختلف متفاوت باشد، زیرا مدل تنها با یک تصویر و بدون اطلاعات چند منظری کار می‌کند. MiDaS قادر به تخمین جزئیات دقیق و بافت‌های ظریف در صحنه‌های پیچیده است و معمولاً قدرت تفکیک خوبی در مرز بین اجسام مختلف ارائه می‌دهد. این مدل، تصاویر RGB را به عنوان ورودی استفاده می‌کند و نیازی به اطلاعات کالیبراسیون دوربین یا داده‌های چند منظری ندارد. همچنین MiDaS به‌گونه‌ای بهینه شده که با سرعت بالا، حتی برای تصاویر بزرگ، نقشه عمق تولید کند که این ویژگی در کاربردهای بلادرنگ مفید است. نقشه‌های عمق خروجی معمولاً مقادیر پیوسته و صافی دارند که برای تولید مدل‌های سه‌بعدی مناسب هستند، هرچند در نواحی با بافت کم ممکن است دقت تخمین پایین‌تر باشد و عمق‌ها ناهموار به نظر برسند.

تولید ابر نقاط

در این گام، برای تبدیل نقشه عمق به ابر نقاط، از مدل دوربین روزنه‌ای (Pinhole Camera) استفاده شده‌است. در این مدل، رابطه بین مختصات پیکسلی تصویر و مختصات سه‌بعدی در فضای دوربین با توجه به فاصله کانونی و نقطه اصلی دوربین تعیین می‌شود (Owen, 1994). فاصله کانونی و نقطه اصلی، به ترتیب، به‌عنوان پارامترهای اصلی دوربین شناخته می‌شوند. مختصات هر پیکسل در تصویر دوبعدی، که با (x, y) مشخص می‌شود و مقدار عمق مرتبط با آن در نقشه عمق، که با (v) نمایش داده می‌شود، به مختصات سه‌بعدی (X, Y, Z) تبدیل می‌گردند. این تبدیل به کمک روابط هندسی زیر انجام می‌شود:

$$\begin{cases} X = \dfrac{v.(x - c_x)}{f} \\ Y = \dfrac{v.(y - c_y)}{f} \\ Z = F(X, Y) \end{cases} \qquad (۱)$$

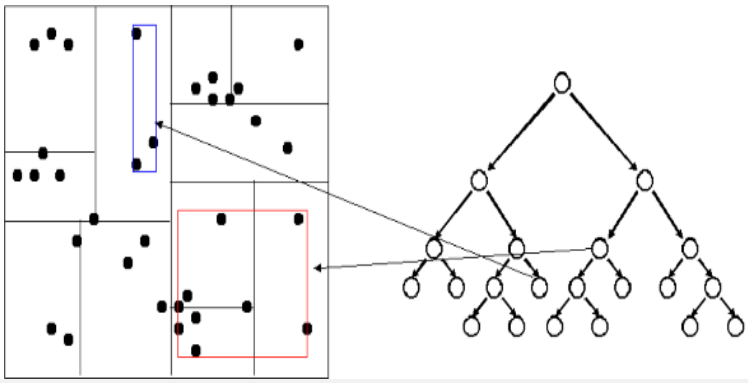
[DOI: 10.22034/KJES.2025.11.1.107622]

در این روابط، (f) فاصله کانونی و (c_x, c_y) مختصات نقطه اصلی تصویر بر روی حسگر دوربین است. با این روش، نقاط دوبعدی تصویر به مختصات سه‌بعدی تبدیل می‌شوند.

پس از تولید ابر نقاط، محاسبه نرمال‌ها با استفاده از الگوریتم‌های مبتنی بر ساختار داده‌ای K-D Tree انجام می‌شود (Zhang Ximin et al., 2013). نرمال‌ها نقش اساسی در دقت و کیفیت بازسازی سه‌بعدی دارند، زیرا اطلاعاتی درباره شیب و جهت سطح ارائه می‌دهند. K-D Tree، با ساختار داده‌ای کارآمد خود، امکان جستجوی سریع و دقیق نزدیک‌ترین همسایه‌ها در فضای چندبعدی را فراهم می‌کند.

در این فرآیند، K-D Tree نقاط همسایه هر نقطه را به سرعت شناسایی کرده و این اطلاعات برای تعیین جهت نرمال سطح در هر نقطه به کار می‌رود. این نرمال‌ها به‌عنوان ورودی در الگوریتم بازسازی سطح پواسون استفاده می‌شوند تا میدان برداری سطح را تعریف کنند.

استفاده از K-D Tree، با کاهش زمان محاسبات و افزایش دقت، امکان محاسبه بهینه نرمال‌های نقاط را فراهم کرده و در نهایت به بهبود کیفیت مدل سه‌بعدی نهایی منجر می‌شود (Häufungsanalyse and Möller, n.d.). در این الگوریتم‌ها، فضا به‌صورت بازگشتی تقسیم می‌شود و هر گره یک فضای کوچک‌تر از نقاط را نگهداری می‌کند. این ساختار به‌ویژه در کاربردهایی مانند گرافیک کامپیوتری، بینایی ماشین و بازسازی سه‌بعدی اهمیت دارد.



شکل ۵- نحوه کارکرد الگوریتم مبتنی بر K-D Tree

Fig. 5. Functioning of the K-D Tree-based algorithm

محاسبه نرمال‌ها برای بهبود دقت بازسازی سطح سه‌بعدی ضروری است.

بازسازی سطح سه‌بعدی

پس از تولید ابر نقاط از نقشه عمق و محاسبه نرمال‌ها، بازسازی سطح انجام می‌شود. این فرآیند به وسیله روش بازسازی سطح پواسون انجام می‌گیرد که یکی از روش‌های مؤثر در بازسازی سطوح از ابر نقاط است.

این روش بر اساس فرضیات پیوستگی و هموار سطح بازسازی‌شده طراحی شده است و با حل معادله دیفرانسیلی پواسون، مشی سه‌بعدی را تولید می‌کند که با نرمال‌های محاسبه‌شده همخوانی دارد. این ویژگی سبب می‌شود که مدل‌های تولیدی دارای جزئیات و دقت بالایی باشند. روش پواسون به‌ویژه برای سطوحی که باید به‌طور طبیعی و پیوسته بازسازی شوند، ایده‌آل است. این روش به خوبی توانایی ایجاد سطوح صاف و یکنواخت را دارد و از آنجا که با توجه به نرمال‌ها عمل می‌کند، می‌تواند جزئیات ظریف را نیز به خوبی حفظ کند. به همین دلیل،

[DOI: 10.22034/KJES.2025.11.1.107622]

استفاده از این روش در فرایند بازسازی سه‌بعدی برای تولید مدل‌های باکیفیت و دقیق بسیار حائز اهمیت است (Kazhdan and Hoppe, 2023).

رابطه پواسون به‌صورت زیر تعریف می‌شود:

$$\nabla.\nabla\phi = \nabla.N \qquad (۲)$$

در رابطه (۲)، (ϕ) تابع پتانسیل سطح، (N) بردار نرمال نقاط، (∇) گرادیان برداری و ($\nabla.$) دیورژانس (divergence) است. دیورژانس، مقدار پراکندگی یا تراکم یک میدان برداری را در یک نقطه مشخص می‌کند(Nolasco et al., 2020).

با حل این معادله، سطح سه‌بعدی به دست می‌آید که به نرمال‌های محاسبه شده از ابر نقاط، منطبق و بسیار نزدیک است و یک مش سه‌بعدی پیوسته و صاف ایجاد می‌کند. این روش به‌ویژه در کاربردهایی که نیاز به بازسازی مدل‌های سه‌بعدی با جزییات دقیق و سطوح هموار دارند، بسیار مؤثر است.

مقایسه مدل سه‌بعدی حاصله از دو روش یادگیری عمیق و فتوگرامتری

برای ارزیابی دقت مدل سه‌بعدی تولیدشده مدل MiDaS، ضروری است آن را با یک مدل سه‌بعدی مبنا و قابل‌اعتماد مقایسه کنیم. به همین منظور، مدل سه‌بعدی شئ موردنظر با استفاده از روش فتوگرامتری تولید شد. در این روش، با ترکیب تصاویر گرفته‌شده از زوایای مختلف، مدلی با جزئیات و دقت بالا بازسازی گردید که به‌عنوان مرجع برای ارزیابی مدل MiDaS به‌کار رفت.

پس از تولید هر دو مدل، اندازه‌گیری‌های نظیر به نظیر بر روی آن‌ها انجام شد تا تفاوت‌های هندسی بین مدل‌ها بررسی شود. این مقایسه شامل اندازه‌گیری فاصله‌ها است و معیاری برای سنجش دقت مدل MiDaS ارائه داد. چنین رویکردی امکان تحلیل عملکرد مدل یادگیری عمیق را در بازسازی سه‌بعدی فراهم می‌کند و تفاوت‌های روش‌های مبتنی بر یادگیری عمیق و فتوگرامتری را برجسته می‌سازد.

برای ارزیابی دقت اندازه‌گیری‌ها، از معیار خطای میانگین کمترین مربعات (RMSE) استفاده شد. این معیار میزان تفاوت میان مقادیر اندازه‌گیری شده بر روی هر دو مدل را محاسبه می‌کند و به‌عنوان شاخصی از دقت بازسازی مدل سه‌بعدی در نظر گرفته می‌شود (Wang et al., 2004). فرمول RMSE در رابطه (۳) آمده است.

$$\mathbf{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_{Multi\_image} - d_{MiDaS})^2} \qquad (۳)$$

در رابطه (۳)، (d_{MiDaS}) و (d_{Multi_image}) به ترتیب نشان‌دهنده طول‌های اندازه‌گیری شده بر روی مدل MiDaS و مدل چند تصویری هستند و (n) تعداد اندازه‌گیری‌هاست.

این مقایسه، این امکان را می‌دهد که نقاط قوت و ضعف مدل یادگیری عمیق را شناسایی کرده که به بهبود الگوریتم‌ها یا معماری مدل موجود منجر می‌شود.

بحث

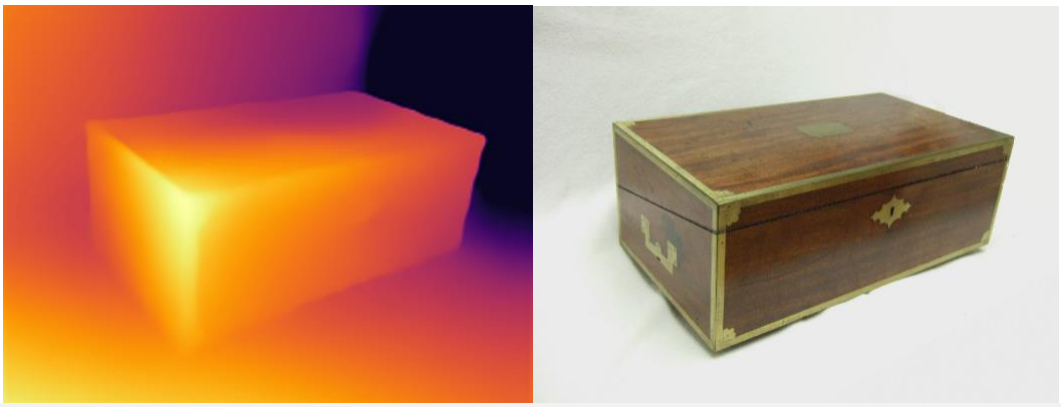
در این پژوهش، ابزارهای مختلفی برای تهیه داده‌ها، بازسازی سه‌بعدی و مقایسه مدل‌ها به کار گرفته شدند:

- مدل یادگیری عمیق MiDaS و تنظیمات مرتبط با معماری آن برای تهیه نقشه عمق و تولید ابر نقاط با استفاده از زبان برنامه‌نویسی پایتون و محیط برنامه‌نویسی Google Colab پیاده‌سازی شد.
- برای بازسازی سطح سه‌بعدی با استفاده از ابر نقاط، از نرم‌افزار CloudCompare بهره گرفته شد.
- تهیه مدل سه‌بعدی به روش فتوگرامتری با استفاده از نرم‌افزار PhotoModeler انجام شد.
- برای مقایسه مدل‌های سه‌بعدی حاصل از دو روش یادگیری عمیق و فتوگرامتری، هر دو مدل در نرم‌افزار CloudCompare بارگذاری و مقایسه شدند.

[DOI: 10.22034/KJES.2025.11.1.107622]

این ابزارها نقش اساسی در مراحل مختلف این پژوهش ایفا کردند و امکان تحلیل دقیق نتایج را فراهم کردند.

یکی از مراحل کلیدی در بازسازی سه‌بعدی با استفاده از مدل MiDaS، تبدیل تصویر RGB به نقشه عمق است. این فرآیند با استفاده از یادگیری عمیق MiDaS انجام شد. شکل (۶)، نقشه عمق تولیدشده توسط این مدل را برای تصویری ورودی نمایش می‌دهد.

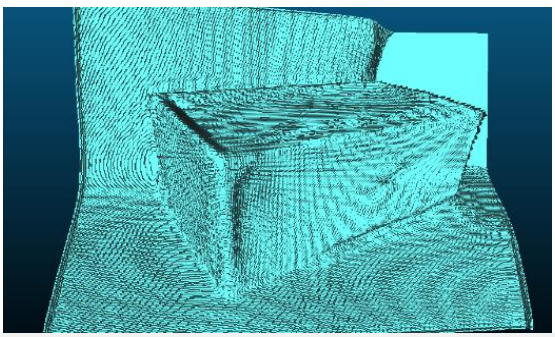


شکل ۶- تصویر RGB و نقشه عمق تولید شده توسط مدل MiDaS

Fig. 6. RGB image and depth map generated by the MiDaS model

همان‌طور که در شکل (۶) مشاهده می‌شود، مدل به‌خوبی توانسته است عمق نسبی سطح شئ را تخمین بزند. مدلMiDaS از ویژگی‌های پیچیده تصاویر و داده‌های آموزشی متنوع برای آموزش استفاده می‌کند تا بتواند عمق نسبی سطوح مختلف را به‌دقت تخمین بزند. در این نقشه عمق، تفاوت‌های شدت رنگ به‌طور دقیق با موقعیت‌های سه‌بعدی اجسام هم‌راستا هستند. بخش‌های نزدیک‌تر به دوربین (مانند وجه جلویی شئ) دارای عمق کمتری هستند و در نتیجه با رنگ‌های روشن‌تری نمایش داده می‌شوند، در حالی که بخش‌های دورتر (مانند وجه بالایی یا کناری) با عمق بیشتر و رنگ‌های تیره‌تری نمایش داده می‌شوند. مدل توانسته است در یک صحنه ساده و بدون تداخل، نقشه عمق پیوسته و همواری ایجاد کند که از دقت و پیوستگی کافی برای استفاده در مراحل بعدی، مانند تولید ابر نقاط و بازسازی سطح سه‌بعدی، برخوردار است.

در گام بعدی، فرایند تبدیل نقشه عمق به ابر نقاط با استفاده از مدل دوربین روزنه‌ای انجام شده است، به‌گونه‌ای که مختصات دوبعدی پیکسل‌ها در نقشه عمق به مختصات سه‌بعدی در فضای دوربین تبدیل می‌شوند. شکل (۷)، ابر نقاط تولیدشده از نقشه عمق، پس از محاسبه نرمال‌ها، نمایش داده شده است.



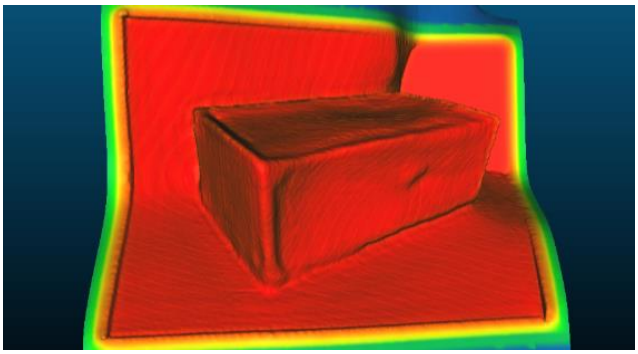
شکل ۷- ابر نقاط بدست آمده از نقشه عمق

Fig. 7. Point cloud obtained from the depth map

[DOI: 10.22034/KJES.2025.11.1.107622]

با توجه به شکل (۷)، ابر نقاط تولیدشده به‌خوبی هندسه کلی شئ را در فضای سه‌بعدی بازتاب می‌دهند. نقاط مربوط به وجوه مختلف شئ به‌صورت واضح تفکیک شده‌اند، که نشان‌دهنده دقت نقشه عمق اولیه و صحت فرآیند تبدیل است و دارای دقت کافی برای استفاده در مرحله بازسازی سطح سه‌بعدی، هستند. این نتیجه نشان‌دهنده عملکرد مناسب مدل MiDaS در ارائه اطلاعات هندسی اولیه از تک تصویر است.

پس از تولید ابر نقاط، مرحله بازسازی سطح سه‌بعدی برای ایجاد یک مدل پیوسته و هموار انجام شد. در این پژوهش، بازسازی سطح با استفاده از الگوریتم پواسون انجام گرفت. این روش، با استفاده از نرمال‌های محاسبه‌شده از ابر نقاط، یک سطح سه‌بعدی پیوسته تولید می‌کند که به‌طور دقیق با هندسه اولیه تطابق دارد. دقت بازسازی به عوامل مختلفی مانند تعداد ابر نقاط، میزان نویز موجود در داده‌ها و پارامترهای استفاده‌شده برای تنظیم الگوریتم پواسون بستگی دارد. به‌طور کلی، اگر ابر نقاط با دقت بالا و نرمال‌ها به درستی محاسبه شوند، این روش می‌تواند مدل سه‌بعدی دقیقی را تولید کند که به هندسه اولیه نزدیک است و تفاوت‌های جزئی میان سطوح مختلف را به خوبی بازسازی می‌کند. شکل (۸)، مش سه‌بعدی حاصل از فرآیند بازسازی سطح پواسون را نمایش می‌دهد.



شکل ۸- مش سه‌بعدی حاصله از روش بازسازی سطح پواسون

Fig. 8. 3D mesh resulting from the Poisson surface reconstruction method

مش سه‌بعدی ارائه‌شده در شکل (۸)، به‌خوبی جزئیات هندسی شئ را بازسازی کرده است. وجوه شئ به‌صورت صاف و پیوسته نمایش داده شده‌اند و مرزها به‌طور دقیق تفکیک شده‌اند. این نشان‌دهنده عملکرد مناسب الگوریتم بازسازی سطح پواسون در تولید مشی هموار و دقیق از ابر نقاط است.

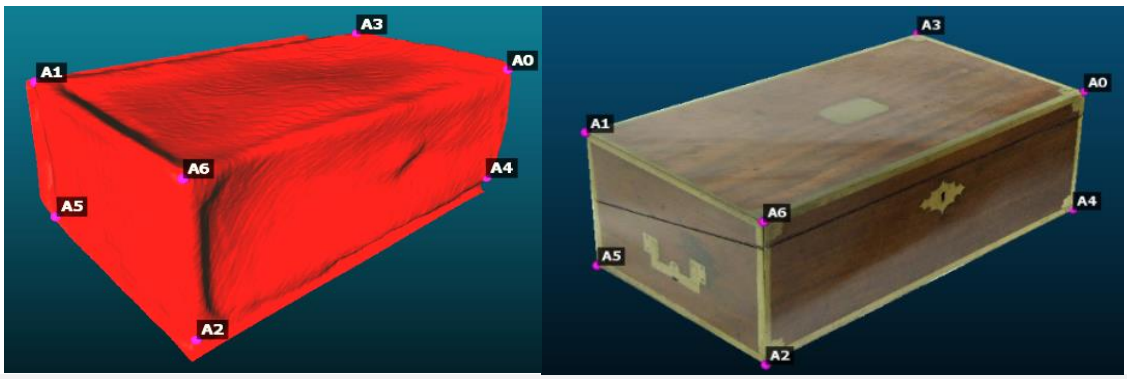
با این حال، ممکن است در بخش‌هایی از مدل، مانند لبه‌ها، جزئیات کمتری مشاهده شود. این مسئله معمولاً به کیفیت نقشه عمق اولیه و تراکم نقاط در مرحله تولید ابر نقاط بستگی دارد. مش سه‌بعدی تولیدشده دارای دقت کافی برای مقایسه با مدل سه‌بعدی حاصل از روش فتوگرامتری است. در شکل (۹)، مدل سه بعدی حاصل از به کارگیری روش فتوگرامتری، نمایش داده شده است.

[DOI: 10.22034/KJES.2025.11.1.107622]

شکل ۹- مدل سه‌بعدی ایجاد شده از روش فتوگرامتری

Fig. 9. 3D model created using the photogrammetry method

این مدل با استفاده از تکنیک فتوگرامتری تولید شده‌است و به‌عنوان مبنایی برای مقایسه با مدل سه‌بعدی روش یادگیری عمیق MiDaS، مورد استفاده قرار می‌گیرد. این مقایسه شامل اندازه‌گیری فاصله بین نقاط متناظر در هر دو مدل است. محاسبه این فاصله‌ها به ارزیابی میزان تطابق هندسی و تشخیص تفاوت‌ها بین دو روش کمک می‌کند. شکل (۱۰)، نقاط متناظر روی هر دو مدل سه‌بعدی را نمایش می‌دهد.



شکل ۱۰- نقاط متناظر بر رو هر دو مدل جهت اندازه‌گیری فواصل

Fig. 10. Corresponding points on both models for distance measurement

برای اندازه‌گیری طول بین رئوس، هر دو مدل سه‌بعدی بارگذاری شده و با استفاده از ابزارهای موجود، این اندازه‌گیری‌ها انجام شد. جدول (۳) نتایج این اندازه‌گیری‌ها را برای هر دو مدل نمایش می‌دهد.

[DOI: 10.22034/KJES.2025.11.1.107622]

جدول ۳- اندازه‌گیری خط‌های مبنای متناظر بر روی مدل MiDaS و مدل فتوگرامتری بر حسب سانتی‌متر

Table 3 - Measurement of baseline distances on the MiDaS model and the photogrammetry model in centimeters

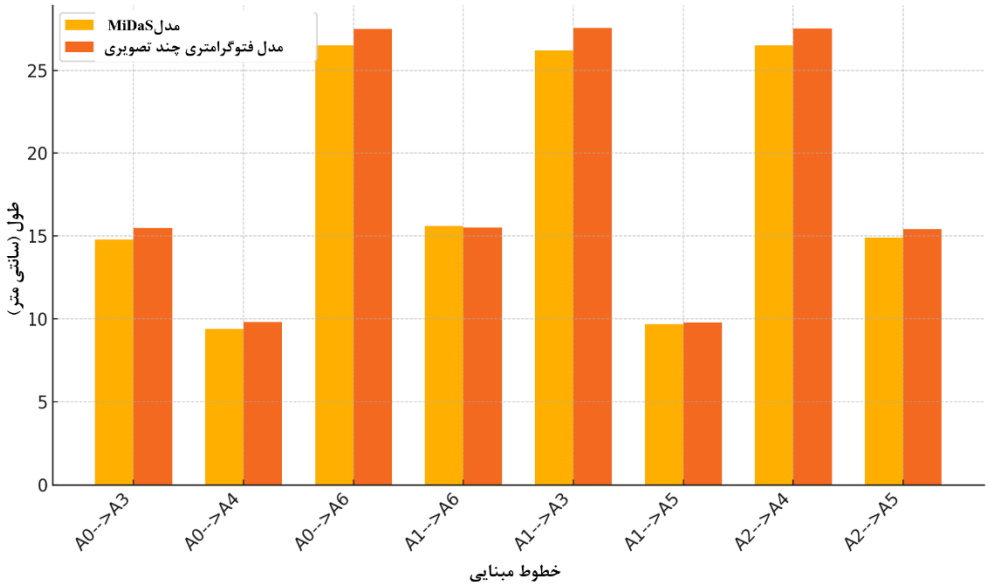
خط مبنا	طول بر روی مدل MiDaS	طول بر روی مدل فتوگرامتری
A0-->A3	۱۴/۸	۱۵/۴۸
A0-->A4	۹/۴	۹/۸۲
A0-->A6	۲۶/۵	۲۷/۵
A1-->A6	۱۵/۶	۱۵/۵۳
A1-->A3	۲۶/۲	۲۷/۵۵
A1-->A5	۹/۷	۹/۷۸
A2-->A4	۲۶/۵	۲۷/۵۲
A2-->A5	۱۴/۹	۱۵/۴۱

نتایج ارائه شده در جدول (۳) نشان می‌دهد که طول‌های اندازه‌گیری شده بر روی مدل MiDaS در اکثر موارد تطابق قابل‌قبولی با مدل مبنا (مدل فتوگرامتری) دارند. اختلافات جزئی مشاهده‌شده در برخی خطوط ممکن است به دلیل محدودیت دقت مدل MiDaS در تخمین عمق نسبی در برخی نواحی باشد.

با استفاده از رابطه (۳)، مقدار RMSE محاسبه‌شده برای مدل سه‌بعدی حاصل از MiDaS برابر با ۰/۷۷۵ سانتی‌متر یا به عبارتی حدود ۸ میلی‌متر است همچنین لازم به ذکر است که میانگین خطای کمترین مربعات برای فواصل اندازه‌گیری شده با استفاده از روش فتوگرامتری در نرم افزار PhotoModeler برابر با ۰/۰۳۷ سانتی‌متر است، که نشان‌دهنده اختلاف نسبی کم (کمتر از یک سانتی‌متر) بین دو مدل است. این مقدار بیانگر آن است که مدل MiDaS، علیرغم استفاده از تنها یک تصویر تک نما، توانسته است بازسازی هندسی مناسبی ارائه دهد. این نتیجه، کاربردپذیری مدل MiDaS در شرایطی که امکان تصویربرداری چند نمایی وجود ندارد را تقویت می‌کند.

شکل (۱۱)، تفاوت طول‌های مبنایی را که بر روی هر دو مدل اندازه‌گیری شده‌اند، نمایش می‌دهد. این نمودار به‌طور بصری کمک می‌کند تا تفاوت‌های موجود بین نتایج دو روش بهتر درک شود.

[DOI: 10.22034/KJES.2025.11.1.107622]



شکل ۱۱- مقایسه طول‌های مبنایی اندازه‌گیری شده بر روی مدل MiDaS و مدل فتوگرامتری

Fig. 11. Comparison of baseline lengths measured on the MiDaS model and the photogrammetry model

این مقدار RMSE به‌طور کلی حاکی از دقت قابل قبول مدل MiDaS در بازسازی سه‌بعدی است، اما می‌تواند به‌طور خاص تحت تأثیر چندین عامل قرار گیرد.

- کیفیت نقشه عمق MiDaS: مدل MiDaS برای تولید نقشه عمق از تک تصویر طراحی شده‌است. با این حال، دقت تخمین عمق به‌ویژه در نواحی با بافت کم یا دور از دوربین ممکن است کاهش یابد. به‌ویژه در نواحی با تغییرات پیچیده هندسی یا شرایط نوری نامناسب، ممکن است منجر به خطاهای بیشتری در نقشه عمق و در نتیجه مدل سه‌بعدی شود که نیاز به بررسی و آزمون‌های متعدد در شرایط مختلف عکس‌برداری دارد. همچنین برای افزایش کیفیت نقشه عمق حاصله، آموزش مدل با تصاویر دارای شرایط متفاوت، می‌تواند تعمیم‌پذیری مدل را افزایش دهد.
- تأثیر تصاویر آموزشی: دقت مدل MiDaS ممکن است به کیفیت و تنوع داده‌های آموزشی آن بستگی داشته باشد. اگر مدل با داده‌های آموزشی متنوعی از اشیاء مشابه آموزش دیده باشد، نتایج بهتری ارائه می‌دهد. عدم تنوع کافی در داده‌های آموزشی می‌تواند بر دقت تخمین عمق تأثیر بگذارد.
- پیش‌پردازش تصویر: مراحل پیش‌پردازش مانند تغییر اندازه و نرمال‌سازی تصویر می‌تواند بر دقت تخمین عمق تأثیر بگذارد. تنظیمات نادرست یا ناکافی در این مراحل ممکن است منجر به نقص‌هایی در نقشه عمق و در نتیجه در مدل سه‌بعدی شود.
- پراکندگی نقاط در فضای سه‌بعدی: روش‌های مختلف برای تبدیل نقشه عمق به ابر نقاط و سپس به مدل سه‌بعدی می‌توانند تأثیر زیادی بر دقت مدل نهایی داشته باشند. استفاده از الگوریتم‌های محاسبه نرمال‌ها و بازسازی سطوح با توجه به ساختار داده‌های ابر نقاط

[DOI: 10.22034/KJES.2025.11.1.107622]

می‌تواند باعث تغییراتی در کیفیت مدل‌های سه‌بعدی شود.

اختلاف‌های مشاهده‌شده در نتایج می‌تواند بر روی کاربردهای بسیار دقیق تأثیر بگذارد. برای کاربردهایی که نیاز به دقت بالا دارند، مانند تجزیه و تحلیل‌های صنعتی، اهمیت دقت بیشتر در تخمین عمق و بازسازی مدل‌های سه‌بعدی غیرقابل انکار است. مدل‌های مبتنی بر MiDaS، با وجود قابلیت‌های پیشرفته خود، ممکن است نیاز به اصلاحات و بهینه‌سازی‌های بیشتری داشته باشند تا بتوانند در تمامی شرایط به دقت مطلوب برسند.

همچنین، نتیجه به‌دست‌آمده از مقایسه مدل‌ها می‌تواند راهنمایی برای انتخاب مناسب‌ترین روش بازسازی سه‌بعدی در کاربردهای خاص باشد. به‌طور کلی، مدل MiDaS به دلیل قابلیت‌های خود در پردازش سریع و استفاده از یک تصویر برای تخمین عمق، می‌تواند گزینه‌ای مناسب برای کاربردهای بلادرنگ باشد، ولی در مواردی که نیاز به دقت بسیار بالا است، نسخه‌های پیچیده‌تر مدل MiDaS ممکن است مزیت‌هایی داشته باشند.

نتیجه‌گیری

در این تحقیق، دقت مدل‌های سه‌بعدی حاصل از نقشه عمق تولیدشده با استفاده از مدل یادگیری عمیق MiDaS و مدل‌های سه‌بعدی به‌دست‌آمده از روش فتوگرامتری چند تصویری مقایسه شد. نتایج به‌دست‌آمده از ارزیابی با استفاده از معیار RMSE، که مقدار آن برابر با ۰/۷۷۵ سانتی‌متر بود، نشان‌دهنده توانمندی مدل MiDaSدر تولید مدل‌های سه‌بعدی با دقت قابل قبول و عملکرد مناسب روش بازسازی سطح پواسون است. این مدل با استفاده از تنها یک تصویر و بدون نیاز به موقعیت و وضعیت ایستگاه‌های تصویربرداری قادر به تولید مدل‌های سه‌بعدی با دقت نسبتاً خوب است. این ویژگی، MiDaS را به گزینه‌ای مناسب برای کاربردهای بلادرنگ و شرایطی که دسترسی به تصاویر چندگانه محدود است، تبدیل می‌کند.

اگرچه استفاده از مدل‌های تخمین عمق مبتنی بر یادگیری عمیق مانند MiDaS امکان استخراج اطلاعات عمق از تصاویر تک نمایی را فراهم کرده است، اما این روش‌ها به‌طور ذاتی با چالش‌هایی مواجه‌اند. از جمله مهم‌ترین چالش‌ها می‌توان به تخمین نادرست عمق در شرایط نوری نامناسب، وجود سطوح همگن یا فاقد بافت، و اشیایی با هندسه پیچیده اشاره کرد. در چنین شرایطی، مدل‌های یادگیری محور به دلیل وابستگی به نشانه‌های بصری مانند سایه، پرسپکتیو، و تقارن، ممکن است تخمین نادقیقی ارائه دهند.

افزون بر این، فرآیند بازسازی سه‌بعدی از روی نقشه عمق نیز به‌تنهایی عاری از خطا نیست. خطاهایی مانند عدم تطابق مقیاس، نویز در داده‌های عمق، و محدودیت روش‌های بازسازی سطح مانند پواسون می‌توانند باعث کاهش دقت مدل نهایی شوند. در تحقیق حاضر، از روش بازسازی سطح پواسون برای تولید مدل سه‌بعدی استفاده شده است. هرچند این روش برای ایجاد سطوح پیوسته کارآمد است، اما در صورت وجود خطا در تخمین نرمال‌ها یا داده‌های ورودی، ممکن است باعث تولید سطوح ناهموار یا نادرست شود.

نتایج مقایسه با مدل فتوگرامتری چند تصویری نیز نشان داد که همچنان اختلاف‌هایی در دقت بین دو مدل وجود دارد. این اختلاف‌ها می‌تواند ناشی از تفاوت در نوع داده‌های ورودی (تصویر تکی در برابر چند تصویر)، و اطلاعات هندسی بیشتر در فتوگرامتری باشد. با این حال، نتایج نشان می‌دهد که با بهینه‌سازی مراحل بازسازی سطح و استفاده از مدل‌های پیشرفته‌تر تخمین عمق، می‌توان این اختلاف را کاهش داد و به دقتی قابل‌قبول برای کاربردهای عملی دست یافت.

یکی از چالش‌های اساسی در تخمین عمق، مربوط به نواحی یکنواخت و فاقد بافت است که ویژگی‌های قابل تشخیص برای الگوریتم وجود ندارد. این مسئله نه تنها در مدل‌های مبتنی بر یادگیری عمیق، بلکه در روش‌های رایج مانند فتوگرامتری برد کوتاه نیز مشاهده می‌شود، چرا که در هر دو روش، استخراج یا تطابق ویژگی در چنین نواحی دشوار است. از جمله راهکارهای پیشنهادی برای کاهش این خطا می‌توان به استفاده از داده‌های مکمل مانند تصویر حرارتی، کانال

[DOI: 10.22034/KJES.2025.11.1.107622]

مادون قرمز یا استفاده از نور ساخت‌یافته در مرحله تصویربرداری اشاره کرد. همچنین استفاده از شبکه‌های پیش آموزش دیده بر روی داده‌های متنوع‌تر و شامل سطوح یکنواخت، یا بهره‌گیری از روش‌های پس پردازش مانند هموارسازی ژئومتریک یا فیلترهای سازگار با مرزهای تصویر(edge-aware filters) ، می‌تواند به بهبود نتایج در این نواحی کمک کند. بررسی اثر این راهکارها بر روی دقت مدل در نواحی کم ویژگی، می‌تواند به‌عنوان موضوعی برای تحقیقات آتی مطرح شود.

با توجه به پیشرفت‌های اخیر در مدل‌های یادگیری عمیق، به‌ویژه نسخه‌های پیچیده‌تر مدل MiDaS، امکان بهبود دقت مدل‌های سه‌بعدی تولیدشده از تک تصویر وجود دارد. پیشنهاد می‌شود که در تحقیقات آینده، با استفاده از نسخه‌های پیشرفته‌تر مدل MiDaS، دقت مدل‌های سه‌بعدی بیشتر افزایش یابد. استفاده از ترکیب روش‌های تک تصویری و چند منظری می‌تواند به بهبود دقت و جزئیات مدل‌های سه‌بعدی کمک کند. این ترکیب می‌تواند از نقاط قوت هر دو روش بهره‌برداری کند و دقت بیشتری را فراهم آورد. استفاده از الگوریتم‌های پیشرفته‌تر برای بازسازی سطح، مانند روش‌های ترکیبی و الگوریتم‌های جدید، می‌تواند به افزایش کیفیت و دقت مدل‌های سه‌بعدی کمک کند.

References

Cazals, F., Giesen, J., 2004. Delaunay Triangulation Based Surface Reconstruction. Ideas and Algorithms. Institut national de recherche en informatique et en automatique 1–45.

Choe, J., Im, S., Rameau, F., Kang, M., Kweon, I.S., 2021. VolumeFusion: Deep Depth Fusion for 3D Scene Reconstruction. Proceedings of the IEEE International Conference on Computer Vision 16066–16075.

Chu, P.M., Sung, Y., Cho, K., 2019. Generative Adversarial Network-Based Method for Transforming Single RGB Image into 3D Point Cloud. IEEE Access 7, 1021–1029.

Eigen, D., Puhrsch, C., Fergus, R., 2014. Depth map prediction from a single image using a multi-scale deep network. Advances in Neural Information Processing Systems 3, 2366–2374.

Eldesokey, A., Felsberg, M., Khan, F.S., 2020. Confidence Propagation through CNNs for Guided Sparse Depth Regression. IEEE Transactions on Pattern Analysis and Machine Intelligence 42, 2423–2436.

Fusiello, A., 2024. Computer Vision: Three-dimensional Reconstruction Techniques. Springer International Publishing Cham.

Häufungsanalyse, C., Möller, P.R., n.d. Non-Standard-Datenbanken und Data Mining Übersicht.

Hermann, M., Ruf, B., Weinmann, M., Hinz, S., 2020. Self-Supervised Learning for Monocular Depth Estimation from Aerial Imagery. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 5, 357–364.

Hristova, H., Abegg, M., Fischer, C., Rehush, N., 2022. Monocular Depth Estimation in Forest Environments. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 43, 1017–1023.

Kazhdan, M., Hoppe, H., 2023. Distributed Poisson Surface Reconstruction. Computer Graphics Forum 42.

Khan, F., Salahuddin, S., Javidnia, H., 2020. Deep learning-based monocular depth estimation methods—a state-of-the-art review. Sensors (Switzerland) 20, 1–16.

Lunscher, N., Zelek, J., 2018. Deep learning whole body point cloud scans from a single depth map. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops 1208–1215.

Ming, Y., Meng, X., Fan, C., Yu, H., 2021. Deep learning for monocular depth estimation: A review. Neurocomputing 438, 14–33

Najaf, M., Arefi, H., Amirkolaee, H.A., Farajelahi, B., 2023. Monocular Depth Estimation of Google Earth Images Using Convolutional Neural Networks. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 10, 589–594.

Nolasco, C., Jácome, N.J., Hurtado-Lugo, N.A., 2020. Applications of the Poisson and diffusion equations to materials science. Journal of Physics: Conference Series 1587.

Owen, T., 1994. Three-Dimensional Computer Vision: A Geometric Viewpoint by Olivier Faugeras The MIT Press, London, UK, ISBN 0–262–06158–9, 1993, 663 pages incl index (£58·50). Robotica 12, 475–475.

Ozden, K.E., Schindler, K., van Gool, L., 2007. Simultaneous Segmentation and 3D Reconstruction of Monocular Image Sequences, in: 2007 IEEE 11th International Conference on Computer Vision. IEEE, 1–8.

Piccinelli, L., Yang, Y.H., Sakaridis, C., Segu, M., Li, S., Gool, L. Van, Yu, F., 2024. UniDepth: Universal Monocular Metric Depth Estimation, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 10106–10116.

Puscas, M.M., Xu, D., Pilzer, A., Sebe, N., 2019. Structured Coupled Generative Adversarial Networks for Unsupervised Monocular Depth Estimation. Proceedings of the 2019 International Conference on 3D Vision (3DV) 18–26.

[DOI: 10.22034/KJES.2025.11.1.107622]

Rajapaksha, U., Sohel, F., Laga, H., Diepeveen, D., Bennamoun, M., 2024. Deep Learning-based Depth Estimation Methods from Monocular Image and Videos: A Comprehensive Survey. ACM Computing Surveys 56, 1–51.

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., Koltun, V., 2020. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer XX, 1–14.

Salzmann, M., Fua, P., 2010. Deformable Surface 3D Reconstruction from Monocular Images. Synthesis Lectures on Computer Vision 2, 1–113.

Saxena, A., Chung, S.H., Ng, A.Y., 2008. 3-D depth reconstruction from a single still image. International Journal of Computer Vision 76, 53–69.

Saxena, A., Sun, M., Ng, A.Y., 2007. 3-D reconstruction from sparse views using monocular vision. Proceedings of the IEEE International Conference on Computer Vision.

Silberman, N., Hoiem, D., Kohli, P., Fergus, R., 2012. Indoor segmentation and support inference from RGBD images. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 7576 LNCS, 746–760.

Wandt, B., Ackermann, H., Rosenhahn, B., 2016. 3D Reconstruction of human motion from monocular image sequences. IEEE Transactions on Pattern Analysis and Machine Intelligence 38, 1505–1516.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13, 600–612.

Welponer, M., Stathopoulou, E.K., Remondino, F., 2022. Monocular Depth Prediction in Photogrammetric Applications. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 43, 469–476.

Zhang Ximin, Liu Ran, Zhou Yiyuan, Wan Wanggen, Lu Libing, 2013. Normal estimation algorithm for point cloud using KD-tree, in: IET International Conference on Smart and Sustainable City 2013 (ICSSC 2013). Institution of Engineering and Technology 286–289.

[DOI: 10.22034/KJES.2025.11.1.107622]