www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 15(3), Sep-Dec 2024, 115-125. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2024.v15i3.521

2D23D: 3D images from 2D images using statistical filters and color shading

Marco Antonio García-Márquez[1,2], *Josue Román Martínez-Mireles*[1,2,†], *Jazmín Rodríguez-Flores* [1,2], *Arturo Austria-Cornejo* [1,2], *Jorge A. Ruiz-Vanoye* [1]

[1] Polytechinc University of Pachuca (UPP),Mexico.
[2] Centro Educativo Tenam, Mexico.

E-mails
† jmartinez@upp.edu.mx

Abstract. Below is the development of an algorithm that allows generating 3D models from 2D images for analysis and visualization, with an application focused on images from the health and space sectors, domains chosen for their contrast characteristics. The characteristics used in the project are presented and determined, as well as 3D programming and visualization environment tools that allow the generation of voxel depth, as well as the management of image formats from the medical or spatial sector, including the results obtained with images. from both sectors, from which the 3D models were created.
Keywords: image, depth, color, mesh, 3D

Article Info
Received Sep 8, 2024
Accepted Sep 28, 2024

1 Introduction

Since the creation of digital images and the use of the pixel proposed by Russel A. Kirch in 1957, as well as the creation of the concept of voxel in three-dimensional space, any object or visible element within space is composed of three dimensions that define its location and are complemented by physical poroperties such as width, height and depth. In this field, a digital image is composed as a two-dimensional representation of a three-dimensional environment, which generates the “direct” loss of depth information, however, they generate a context where techniques are implemented that allow the analysis and two-dimensional representations for the study of physical phenomena. It is a fact that images have a color distribution that can be represented through its histogram, which can be equalized, expanded or specified as presented in (García et al., 2014), a work that proposes to improve the contrast through the local adjustment of the histogram by applying diffusions that adapt the intensity of colors contained in each image either by equalization, expansion or specification of a histogram searched.

The creation of 3D models from 2D images, seeks to contribute or strengthen the state of the art where image processing research is developed, focused on the generation of solutions based on 3D models, which have a wide range of applications, An example of this is in the medical sector (through images from imaging studies) mentioned by (Fallavollita, 2009) or in the case of space exploration (derived from photographs of specialized telescopes), robotics (for the analysis of the environment in digital images, and the management of virtual reality applications, augmented reality and extended reality (which make extensive use of two-dimensional graphical representations of the real world). (Yang, 2023). To establish and identify the depth property in a real environment, there are schemes based on ultrasonic technology, laser and/or stereo vision, which require more resources, both economic and high performance processing for its implementation in engineering solutions, as an alternative we seek to develop through image processing, functional algorithms for depth recognition in two-dimensional images for high-impact solutions, mainly based on the basic characteristics contained in the image itself. The basic characteristics that can be extracted from an image and that intervene in the depth are: Hue, allows to differentiate each color, in other words, it is the tint or value of each pixel in each channel; Brightness, is the amount of light emitted or reflected by a pixel, that is, its lightness or darkness, Saturation: is the intensity or degree of purity of each color and Contrast: which allows to measure the hue, brightness and saturation with respect to a set of pixels. (Kudo, 2013) (Koschan and Abidi, 2008) (Tavares and Jorge, 2008). When working with three-dimensional representations, the concept of a voxel arises, which is the basic cubic unit that composes them.

From the determination of voxels, the width and height dimensions can be complemented for the establishment of depth in the real world.

One of the main difficulties in creating 3D models from 2D images is the estimation of voxel depth. There are several algorithms for estimating voxel depth from 2D images. These algorithms are based on different techniques, such as triangulation, segmentation and machine learning. The creation of three-dimensional content for virtual environments, augmented reality or the implementation of some artificial vision techniques such as texture detection and specific features of scenes is a task that becomes laborious, coupled with the time that depends largely on experts in 2D design and of course in 3D, as well as the tools and techniques they use. The conversion of 2D images to 3D will involve the transformation of two-dimensional information into three-dimensional information from a common digital camera, which complicates the creation of 3D volume models from a single 2D image.

Voxels are the basic elements of 3D models. Their depth determines the distance from the camera to the object they represent. Currently there are several algorithms to estimate voxel depth from 2D images. These algorithms are based on different techniques, such as triangulation, segmentation and machine learning. When using image processing methods, for depth property identification, there are schemes based on ultrasonic, laser and/or stereo vision technology, which require more resources for their implementation in engineering solutions. Studies have been conducted focused on the analysis of the properties of 2D images focused on identifying their impact and correlation with the depth property of the elements captured in the image, identifying the possibility of using the luminosity present in the image, for this reason, this paper focuses on the main problem of determining the depth of the voxels of 3D models from 2D images with a common digital camera, as well as the structuring of information to propose the rendering of the generated model.

2. Main techniques for 2D to 3D conversion

The community has implemented different strategies to obtain three-dimensional information from that contained in the two-dimensional images, the main strategies implemented are:

- *Deep Learning*: Deep neural networks, such as generative adversarial networks (GANs) and convolutional neural networks (CNNs), have revolutionized 2D-to-3D image processing. These techniques enable the generation of realistic 3D models from 2D images and have driven advances in the conversion of 2D portraits to good quality 3D models.
- *Digital Photogrammetry*: Photogrammetry has evolved with the use of advanced digital cameras and processing software. Accurate 3D models can be generated from multiple 2D images through key point matching and triangulation.
- *Augmented and Virtual Reality*: 2D to 3D image conversion plays a key role in augmented and virtual reality applications. This makes it possible to superimpose virtual objects on the real world or to create immersive virtual worlds based on 2D images of the environment.
- *3D Scanning with Mobile Devices*: The increasing availability of advanced sensors in mobile devices has enabled users to scan 3D objects or environments from 2D images, which has applications in design, 3D printing and archaeology, among other fields.
- *3D printing*: 2D to 3D image conversion is used in 3D printing to create high-quality three-dimensional models. 2D drawings can be converted into physical 3D objects with precise details.

2.1 Related Works

The work (Bahmani et al., 2023) describes the process for the generation of 3D scenes from images with furniture scenes distributed in a 2D plane and presents a functional algorithm such as the one proposed in this project as a research proposal. In (Chung, Lee, et al., 2023) a way of "smoothing" the 3D reconstruction of a two-dimensional image by extracting the depth feature of the image itself is presented. Something similar is presented in (Höllein et al., 2023) with a strategy for the generation of 3D scenes from the information provided by a 2D image which incorporates generative text information, which proposes a method to generate scaled and inhabited 3D textured meshes from a text message as input, the model employs monocular depth estimation with a text-conditioned image model and generative algorithms.

A different strategy for the generation of 3D structures from images is in (Quintero and Duque, 2008) where the reconstruction is performed from a set of shots, which are integrated from different perspectives for the creation of 3D objects, similar to a 3D laser scanner. Similarly, (Palos Cuesta, 2012) shows the processing of an image, and through the extraction of features and structures, the 3D reconstruction of an image from a 2D source is performed. Regarding medical applications, (Luna et al., 2020) develops the mechanism for the reconstruction of 3D images from 2D images, in particular the case study is tomography scans of cancerous tumors. By contrasting the results of different algorithms in generating 3D objects from 2D images (Xu et al., 2022), shows the results of the new algorithm, called Nauralift by the authors, with the results of similar algorithms and demonstrates" cleaner" reconstructions without information non-existent. important. These are just some related research works; more related research can be found in the references as it is a highly applicable topic with emphasis on medical applications.

As described above, voxels are the basic element of 3D models its depth determines the distance from the camera to the object they represent (Sella et al., 2023). Currently, there are several algorithms to estimate voxel depth from 2D images. These algorithms are based on different techniques, such as triangulation, segmentation and machine learning. When image processing methods are used, for depth property identification, there are schemes based on ultrasonic, laser and/or stereoscopic vision technology, which require a greater number of resources for their implementation with applied engineering solutions. Studies have been carried out dedicated to the analysis of properties of 2D images and focused on identifying their impact and correlation with the depth property of the elements that make up the image, with the possibility of using the energy reflected in the image itself. This thesis project will address the main problem of determining the depth of the voxels of 3D models from 2D images with a common digital camera, as well as the structuring of the information proposed for the rendering of the generated model (Chung, Oh, and Lee, 2023) (Feng and Singhal, 2023) (Miyanishi et al., 2023) (Jeong et al., 2022) (Ryu et al., 2023)

One of the tools currently available for managing 2D images and creating a 3D model is Open3D proposed by (Zhou et al., 2018) some results are presented in figure 1, this library allows implementing various algorithms and data structures for this purpose. On the other hand, (Tang et al., 2023) presents a proposal where it uses an image to estimate the geometry in two stages, the first that optimizes a neural radiation field and incorporates restrictions from the same image of the front view with previous diffusions of the possible views, and in the second stage transforms the model into textured point clouds and increases realism by taking advantage of the high-quality textures of the referred image. The work (Richardson et al., 2023) proposes to generate 3D textures by dynamically defining a trimap partition of the rendered image with three progression states and presents a diffusion sampling process that uses the trimap representation to generate the textures from different views. Secondly, to add new points to the 3D scene, an alignment algorithm is proposed with the parts of the generated 3D scenes (Chung, Lee, et al., 2023). In (Ranftl et al., 2020) a robust depth and range scale invariant objective is proposed, using multi-objective learning, combining data from different sources, and training image domain encoders. Also, at (Ranftl et al., 2021) they present what they call dense vision transformers, an architecture that leverages vision transformers instead of convolutional networks as the backbone for dense prediction tasks, where features from multiple stages are concentrated during the transformation and images with various resolutions are combined. progressively. them and employ full-resolution prediction using a convolutional decoder, which provides finer and more globally consistent predictions compared to fully convolutional networks.

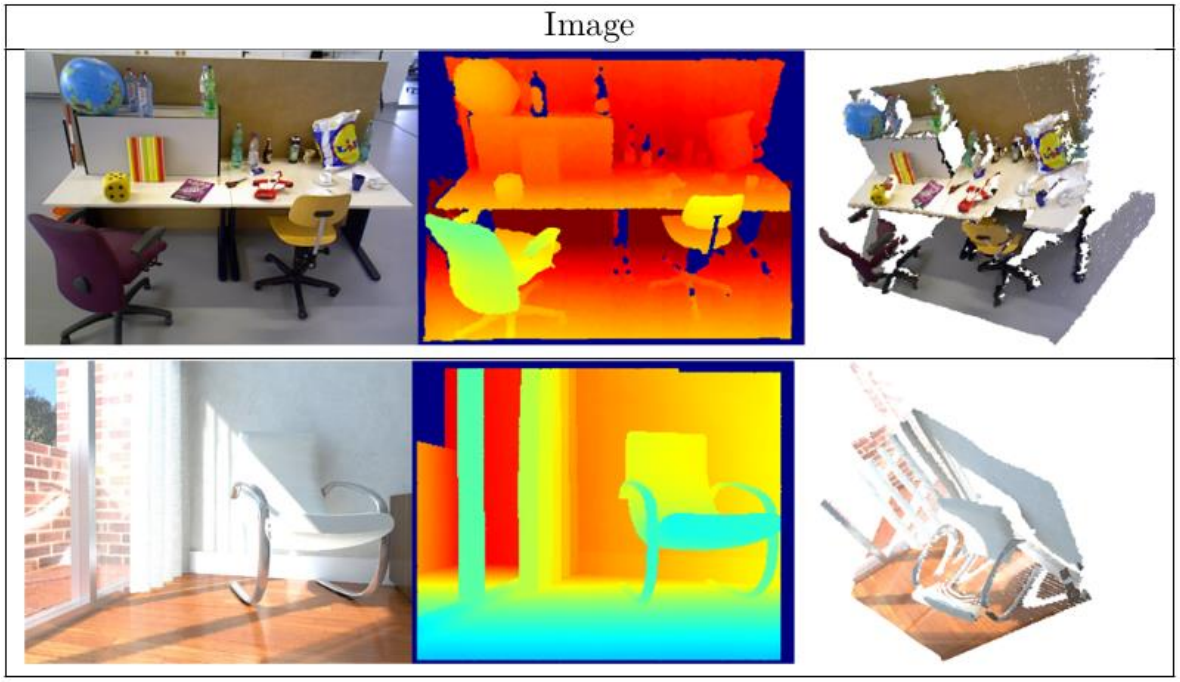

Fig. 1. Demonstration of the Open3D library. (Zhou et al., 2018)

On the other hand, and with the rise in the use of platforms that propose the use of artificial intelligence, as is the case of CSM, 2023, platforms are proposed that are oriented towards the creation or pre-construction of 3D models mainly for video games. virtual reality or augmented reality. On this platform, some images of the study domain were sent, the platform receives the 2D image to convert it into a 3D object, proposes image projections as in (Tang et al., 2023) or (Chung, Lee, et al., 2023) and the results can be seen in Figure 2. The images in Figure 2 are images sent to the CSM platform, original images of the planet Saturn, a thorax and a human eye, taken from the Internet.

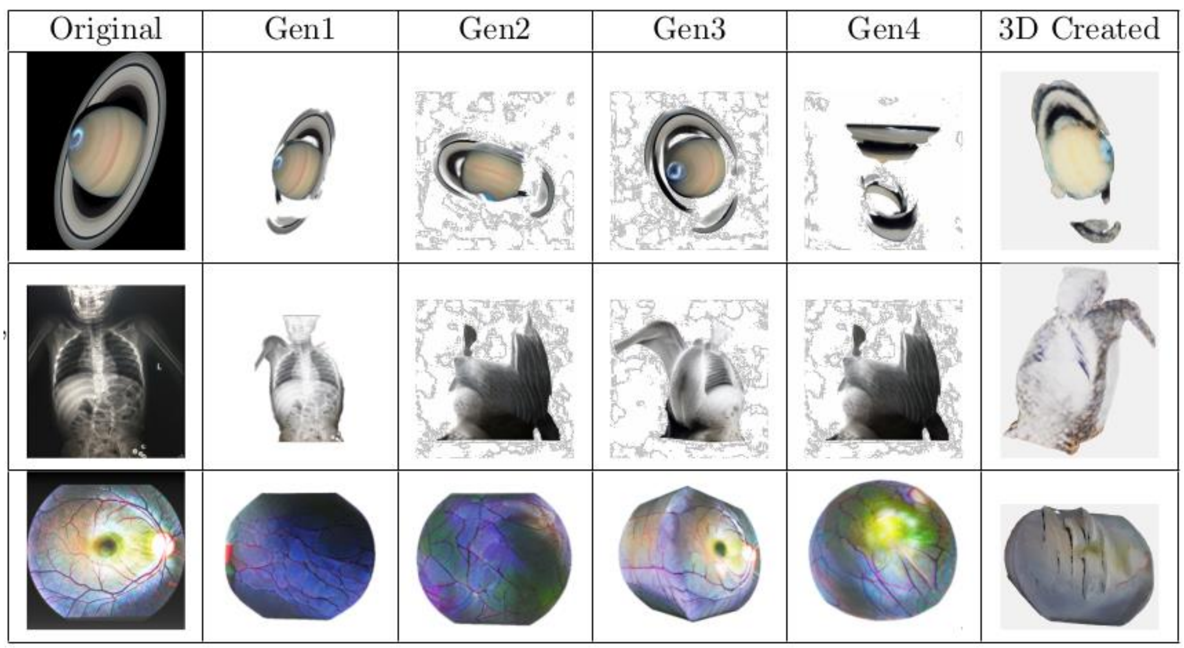

Fig. 2. 3D models generated with CSM generative AI. CSM, 2023

2 Methodology and experimental procedures

The developed methodology and the sections that compose it are presented below with the intention of structuring its development, development of the proposal, justification, contribution, implementation, tests and results. 2D to 3D image processing has advanced significantly thanks to deep learning technology and techniques. This has led to more precise and versatile applications in fields such as medicine, video games, virtual and augmented reality and 3D printing, which continues to drive the development of increasingly sophisticated tools and solutions in this exciting field.
The main problems when generating 3D models with digital images in the medical sector are the following (Sangwine and Horne, 1998) (Koschan and Abidi, 2008):

- Noise: Medical images often have a high level of noise, which can make it difficult to extract accurate information.
- Distortion: Medical images can be distorted by factors such as patient position, imaging equipment, or the image acquisition process.
- Incompleteness: Medical images are often incomplete, capturing only part of the object being viewed.
- Variability: Different types of medical images have unique characteristics and limitations, which can make it difficult to develop 3D model generation methods that are applicable to all cases.

These issues can make it difficult to generate accurate and reliable 3D models of medical objects. To overcome these limitations, several image processing and machine learning techniques have been developed.

- Noise: Noise can be reduced using image processing techniques such as filtering, smoothing, and denoising.
- Distortion: Distortion can be corrected through image registration and calibration techniques.
- Incompleteness: Incompleteness can be addressed through inference and reconstruction techniques.
- Variability: Variability can be addressed using machine learning techniques, such as deep learning.

Despite the challenges, generating 3D models from medical images has great potential to improve medical diagnosis and treatment. 3D models can be used to visualize medical objects in a more realistic and accurate way, which can help better understand the anatomy and pathology of patients. Additionally, 3D models can be used to plan surgeries, perform simulations, and provide information to patients. In recent years there have been important advances in the development of techniques for generating 3D models from medical images. In the case of 3D model incompleteness techniques, advances have made it possible to generate more precise and reliable 3D models, which has led to greater acceptance of this technology in the medical field. With the above, the proposed methodology to generate depth in 2D images is presented.

Some of the characteristics that a 2D image contains, whether locally in each pixel, in a set of pixels or even in the entire image, is the calculation of brightness in equation 3, contrast in equation 3 and saturation with respect to the tone or color of each pixel or set of pixels, these characteristics are part of this proposal.

Compute local and global image brightness in parallel with the following metrics and protect them using maps and ideally with tensors.

$$bright = \frac{1}{N}\sum_{i=0}^{N} x_{i_R} + x_{i_G} + x_{i_B} \tag{1}$$

Calculate and review the local and global brightness of the image.

$$contrast = \max x_{i_{(R+G+B)}} - \min x_{i_{(R+G+B)}} \tag{2}$$

Calculate and review local and global saturation of the image.

$$saturation = \frac{\max x_{i_{(R+G+B)}} - \min x_{i_{(R+G+B)}}}{N} \tag{3}$$

Calculate the local depth of the voxels and the global depth of the image considering the average intensity of a neighborhood, e.g. 3 ∗ 3 and may be larger, considering the energy reflected in the scene and image domain.

$$depth = \frac{1}{C \cdot N}\sum_{j=1}^{C}\sum_{i=1}^{N} x_{ij} \tag{4}$$

Determine the classification of objects with respect to local and global depth.

$$\mu_\beta\left(depth_{x_i}\right) = \left\{newmesh\ if\ depth_{x_i} \geq depthGlobal,\ bind\ if\ depth_{x_i} < depthGlobal\right\} \tag{5}$$

Where newmesh is the closure of the object mesh and opening of the new mesh, while bind is the union of voxels to integrate elements of the 3D model of equation 3. The procedure is shown below with a digital image of the ocular health domain shown in Figure 3, to which the methodology is applied in two aspects, one with the normal process and the other applying local contrast enhancement preprocessing, in both cases, a 3D model is obtained by generating the voxels using the proposed methodology. In these images you can see the objects in the scene according to the calculated characteristics with which the depth of each voxel in the image was created.

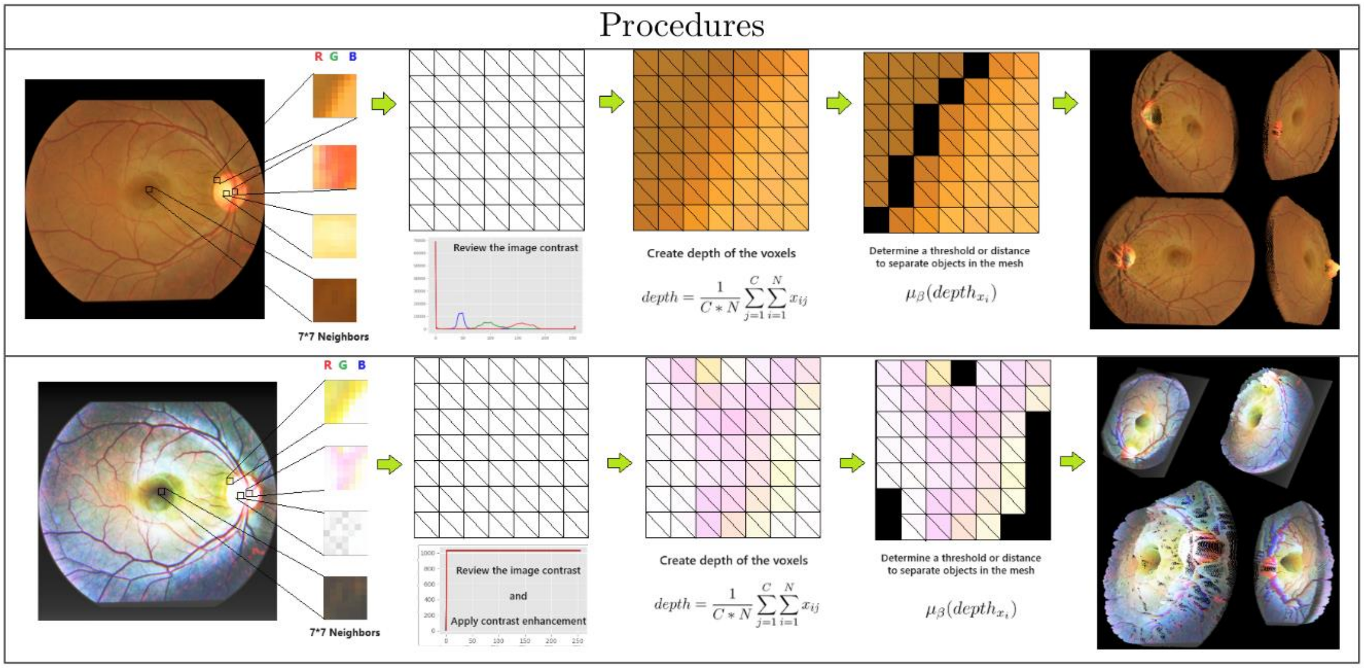

Fig. 3. Procedures that implement the methodology, source image from (Oftalmología, 2021).

The procedures presented correspond to the phases established in the equations to calculate the characteristics of brightness, contrast, saturation and of course the statistical calculation of the depth through a neighborhood that can range from $3 * 3$ and up to $7 * 7$ or $m * n$, considering that it will depend on the depth distribution of each object contained in the image and of course its scope. With the methodology presented, various tests have been satisfactorily carried out and have given some results that will allow us to continue, optimize or expand this research work. In the following section, some of the tests obtained so far will be presented, which were initially developed in the Java languages (Oracle, 1995) through Processing (Fry and Casey Reas, 2001) and Python (Foundation, 2001) with various libraries for handling and image viewing.

3 Results

The figure 4 show a human eye with glaucoma problems, this condition is characteristic when the optic nerve is damaged. As this nerve progressively deteriorates, blind spots in vision (dark part) and increased pressure appear in the eye. As presented in the figure, the first results of the depth calculation, both to the original image and when applying local or global contrast enhancement, can observe the creation of depth without texturing and identify what stands out in The scene includes elements of interest such as nerves, arteries and of course in this case, the damage caused by glaucoma, considering that it allows an expert to identify what is relevant now with a 3D model. Figure 5 shows the generation of depth in a digital image from a fluoroscopy of the digestive system.

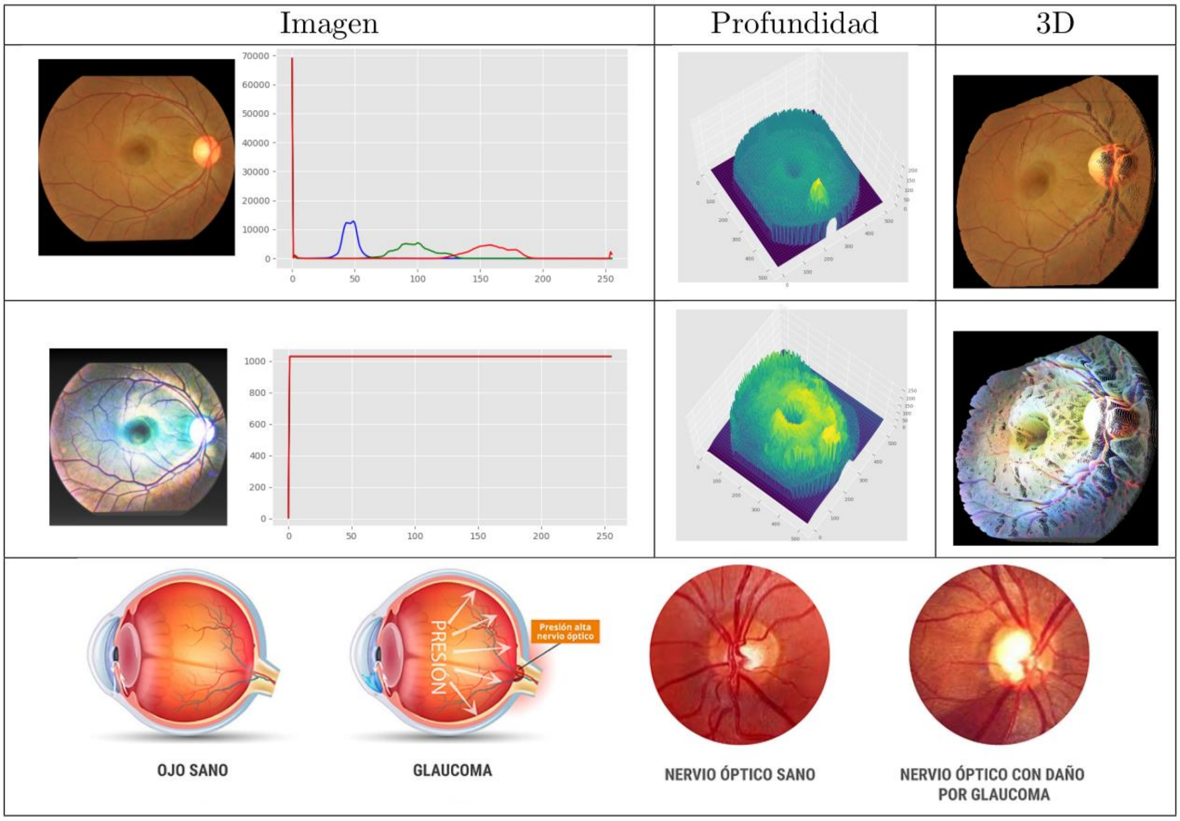

Fig. 4. Human eye with glaucoma, original image and another with local contrast enhancement, prior to 3D generation. Below, the characteristics of glaucoma in the human eye. (Oftalmología, 2021.)

Figures 6 to 8 show the results of applying the algorithm to medical images, depending on the clarity of the image and the reconstructed information, the 3D model can be complemented. The results of radiographs and MRIs are observed, with the generated models the possibility of contributing to medical diagnostic processes through the 3D model is projected.

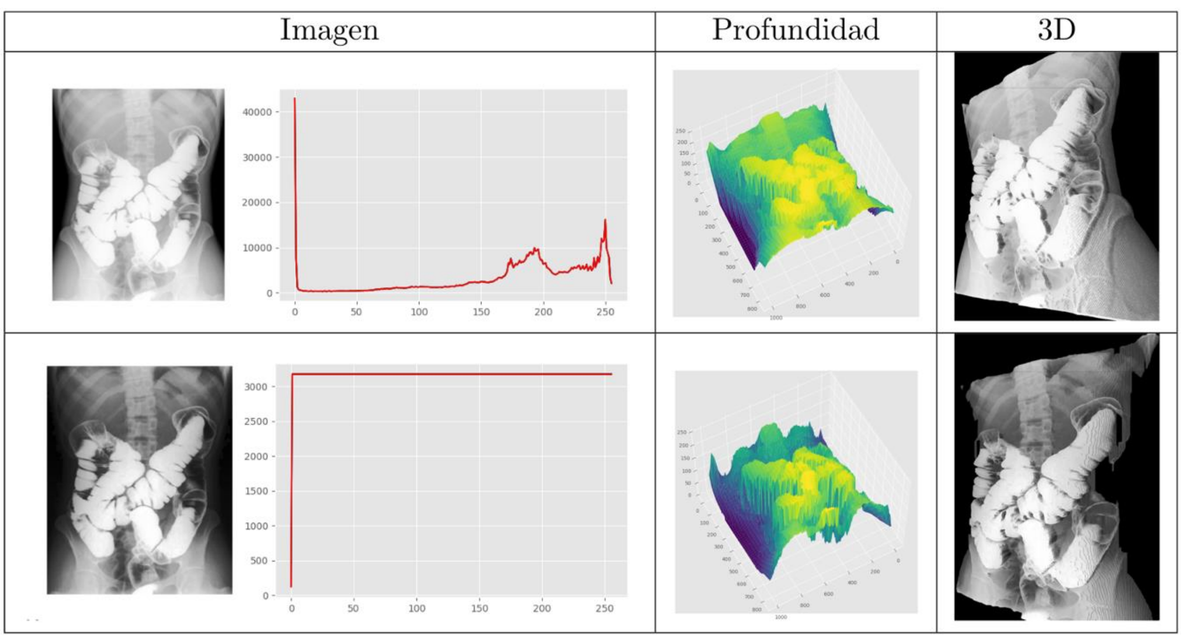

Fig. 5. Digestive system fluoroscopy

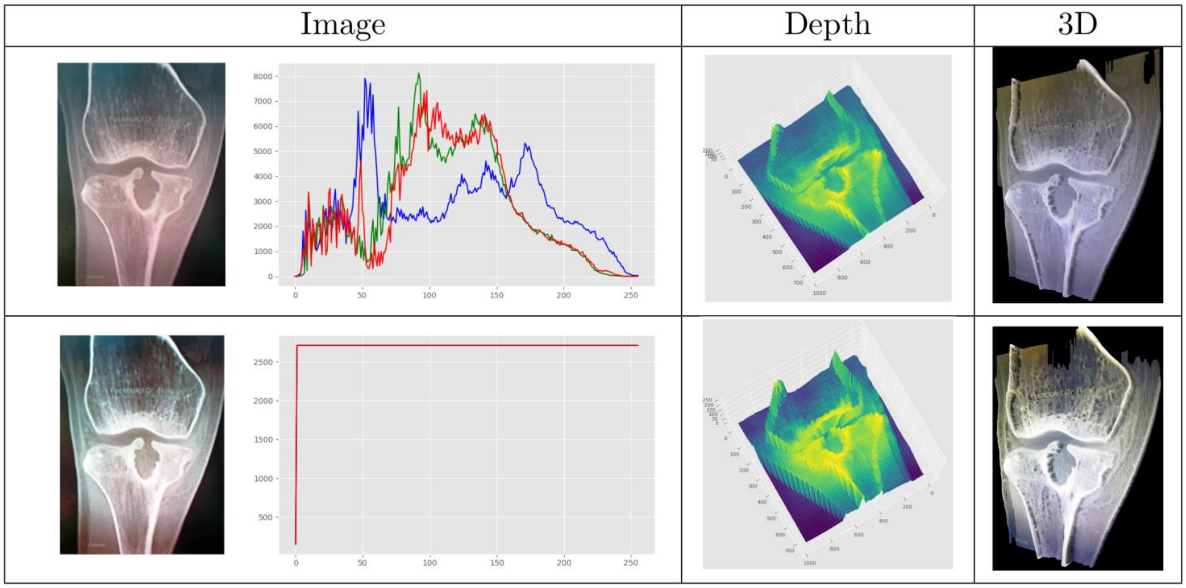

Fig. 6. X-ray of a knee

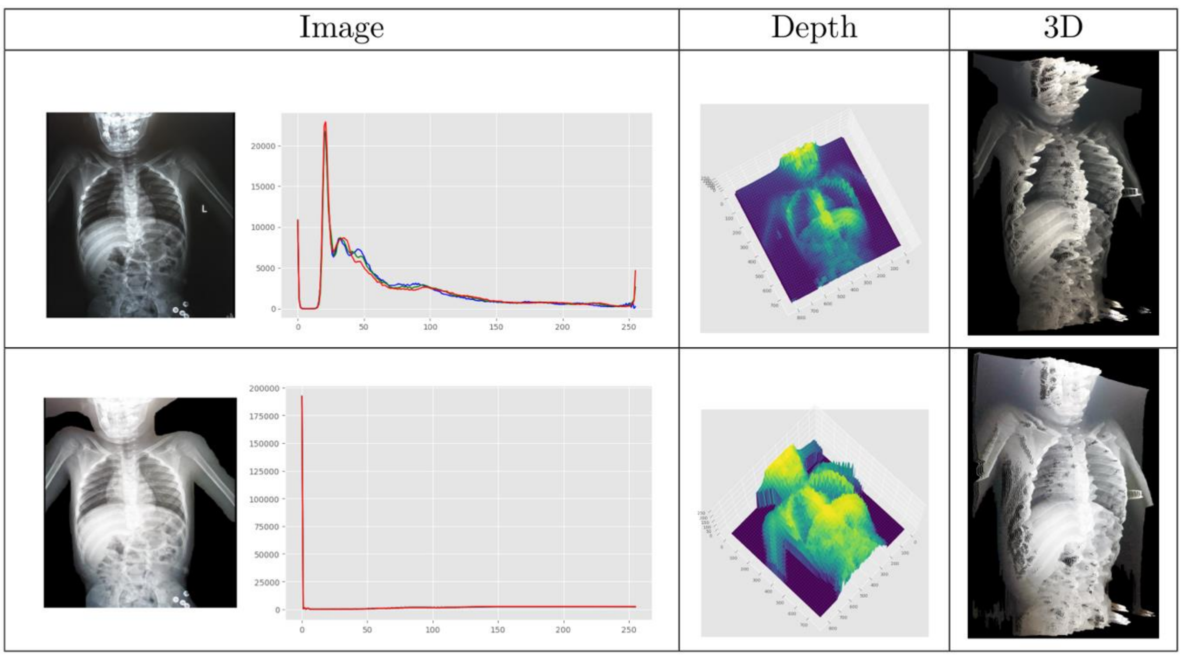

Fig. 7. Chest X-ray

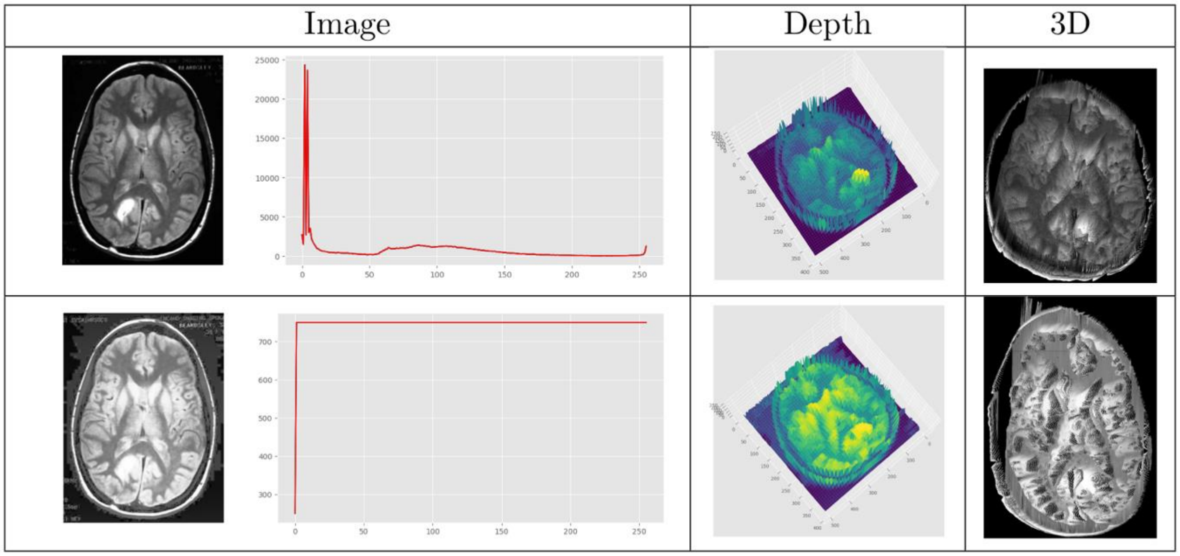

Fig. 8. Partial neural resonance

4 Conclusions

With the above, and considering the progress of the research proposal presented, the development stage of the project and the respective tests with some platforms from the literature reviewed so far have begun. With this first stage, and with the reviewed bibliography, as well as the first tests with algorithms and literature libraries, as well as the proposed algorithm and the results obtained so far, the contribution to the state of the art is considered relevant. In general, it is concluded that it is very likely to achieve the proposed objective of this project with the initial review of the research background and the progress of the presented prototype. Likewise, it was possible to identify the impact of the research project, the tools and main opportunities that will be strengthened during its optimization and development.

With the results shown so far, it is concluded that the tests carried out have been important and significant; some variants have been detected that should be added to the proposal, such as strengthening the preprocessing and thresholding functions of elements that will participate in the definition of depth, the function of separating objects from the scene and consolidatingthe function of generalizing the definition. of depth and total management through defined tensors or objects, with which it is possible to integrate the rest of the characteristics that allow optimizing the render and its deployment on standard visualization platforms. Likewise, it is relevant to comment that the image generation or processing time is acceptable compared to the reviewed literature, besides, the programming languages used so far have been Java and Python, however, other alternatives will be sought that allow an application aimed at users in general and consider interactivity with said platform for direct testing and optimization through experiences. of user.

On the other hand, a data set (2D X-ray images) was identified that contains training images and test images, which integrate images of healthy patients and images of patients with pneumonia and that are used to train a neural network with said set of images and be able to diagnose pneumonia in new images, having said the above and as further work, it is intended to apply said methodology to said set of images to know the behavior of the neural network using 3D models and compare the results to be able to optimize and validate the proposal. Without a doubt, I consider the importance of now being able to redefine the management of 3D digital models or images that allow us to capture the greatest amount of information regarding the three-dimensional or spatial space that surrounds us.

The application of elements of artificial intelligence where prior knowledge about the environments can be added, as well as previous results added to reinforcement learning, is seen as one of the methodologies that can most successfully generate the reconstruction of 3D models. It is expected that depth reconstruction from digital images of the health sector will contribute to the science of medical diagnosis. from providing more information to treating doctors to studying the evolution of patients in internal diseases of the body, whose inspection requires specialized medical equipment, and thus contributing to improving the health of the population.

References

Bahmani, S., Park, J. J., Paschalidou, D., Yan, X., Wetzstein, G., Guibas, L. J., & Tagliasacchi, A. (2023). CC3D: Layout-conditioned generation of compositional 3D scenes. *CoRR, abs/2303.12074*. https://doi.org/10.48550/ARXIV.2303.12074

Chung, J., Lee, S., Nam, H., Lee, J., & Lee, K. M. (2023). Luciddreamer: Domain-free generation of 3D Gaussian splatting scenes. *CoRR, abs/2311.13384*. https://doi.org/10.48550/ARXIV.2311.13384

Chung, J., Oh, J., & Lee, K. M. (2023). Depth-regularized optimization for 3D Gaussian splatting in few-shot images. *CoRR, abs/2311.13398*. https://doi.org/10.48550/ARXIV.2311.13398

CSM. (2023). *CSM: Modelos 3D generativos*. https://3d.csm.ai/

Fallavollita, P. (2009). 3D/2D registration of mapping catheter images for arrhythmia interventional assistance. *CoRR, abs/0910.1844*. http://arxiv.org/abs/0910.1844

Feng, J., & Singhal, P. (2023). 3D face style transfer with a hybrid solution of NeRF and mesh rasterization. *CoRR, abs/2311.13168*. https://doi.org/10.48550/ARXIV.2311.13168

Foundation, P. S. (2001). *Python*. https://www.python.org/

Fry, B., & Casey Reas, P. (2001). *Processing*. https://processing.org/

García, M. A., Ben-Youssef, C., & C., R. S. (2014). Histogram equalization based on the fuzzy local fittingness. In *Decision making and soft computing* (pp. 418–423). https://doi.org/10.1142/9789814619998_0070

Höllein, L., Cao, A., Owens, A., Johnson, J., & Nießner, M. (2023). Text2room: Extracting textured 3D meshes from 2D text-to-image models. *CoRR, abs/2303.11989*. https://doi.org/10.48550/ARXIV.2303.11989

Jeong, Y., Shin, S., Lee, J., Choy, C. B., Anandkumar, A., Cho, M., & Park, J. (2022). Perception: Perception using radiance fields. *CoRR, abs/2208.11537*. https://doi.org/10.48550/ARXIV.2208.11537

Koschan, A., & Abidi, M. (2008). *Digital color image processing*. Wiley. https://books.google.com.mx/books?id=SlXgTyQ86VsC

Kudo, M. (2013). *Contrast harmonic imaging in the diagnosis and treatment of hepatic tumors*. Springer Japan. https://books.google.com.mx/books?id=8HbmCAAAQBAJ

Luna, J. Z. H., et al. (2020). *Procesamiento de imágenes digitales: 2D a 3D aplicado a tumores cancerígenos*. http://repositorio.ugto.mx/handle/20.500.12059/2156

Miyanishi, T., Azuma, D., Kurita, S., & Kawanabe, M. (2023). Cross3DVG: Baseline and dataset for cross-dataset 3D visual grounding on different RGB-D scans. *CoRR, abs/2305.13876*. https://doi.org/10.48550/ARXIV.2305.13876

Oftalmología, G. (2021). *Glaucoma*. https://www.geof.com.ar/claucoma

Oracle. (1995). *Java*. https://www.java.com/es/

Palos Cuesta, J. (2012). *Desarrollo de un procedimiento para la creación de imágenes 3D a partir de imágenes 2D* [Doctoral dissertation, Universitat Politècnica de València].

Quintero, J. F. O., & Duque, O. E. S. (2008). *Generación de estructuras 3D a partir de imágenes 2D* [Doctoral dissertation, Universidad EAFIT, Departamento de Ingeniería de Sistemas].

Ranftl, R., Bochkovskiy, A., & Koltun, V. (2021). Vision transformers for dense prediction. https://doi.org/10.48550/arXiv.2103.13413

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K., & Koltun, V. (2020). Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. https://arxiv.org/abs/1907.01341v3

Richardson, E., Metzer, G., Alaluf, Y., Giryes, R., & Cohen-Or, D. (2023). Texture: Text-guided texturing of 3D shapes. *CoRR, abs/2302.01721*. https://doi.org/10.48550/ARXIV.2302.01721

Ryu, N., Gong, M., Kim, G., Lee, J., & Cho, S. (2023). 360° reconstruction from a single image using space carved outpainting. *CoRR, abs/2309.10279*. https://doi.org/10.48550/ARXIV.2309.10279

Sangwine, S., & Horne, R. (1998). *The colour image processing handbook*. Springer US. https://books.google.com.mx/books?id=oEsZiCt5VOAC

Sella, E., Fiebelman, G., Hedman, P., & Averbuch-Elor, H. (2023). Vox-e: Text-guided voxel editing of 3D objects. *CoRR, abs/2303.12048*. https://doi.org/10.48550/ARXIV.2303.12048

Tang, J., Wang, T., Zhang, B., Zhang, T., Yi, R., Ma, L., & Chen, D. (2023). Make-it-3D: High-fidelity 3D creation from a single image with diffusion prior. *CoRR, abs/2303.14184*. https://doi.org/10.48550/ARXIV.2303.14184

Tavares, J., & Jorge, R. (2008). *Advances in computational vision and medical image processing: Methods and applications*. Springer Netherlands. https://books.google.com.mx/books?id=JykChzLVlogC

Xu, D., Jiang, Y., Wang, P., Fan, Z., Wang, Y., & Wang, Z. (2022). NeuralLift-360: Lifting an in-the-wild 2D photo to a 3D object with 360° views. *CoRR, abs/2211.16431*. https://doi.org/10.48550/ARXIV.2211.16431

Yang, H. (2023). Sketch2CADscript: 3D scene reconstruction from 2D sketch using visual transformer and rhino grasshopper. *CoRR, abs/2309.16850*. https://doi.org/10.48550/ARXIV.2309.16850

Zhou, Q., Park, J., & Koltun, V. (2018). Open3D: A modern library for 3D data processing. *CoRR, abs/1801.09847*. http://arxiv.org/abs/1801.09847