Automated Generation of Tactile Representations of Buildings and Monuments from Pictures: Advanced AI Pipeline Development

Abdulmohsen Ahmed Alghamdi
King Fahd University of Petroleum and Minerals
Information and Computer Science Department
S202171250@kfupm.edu.sa

Project Mentor 1:
Carlo Fonda
Manager, ICTP Scientific Fabrication Laboratory
International Centre for Theoretical Physics (ICTP)
cfonda@ictp.it

Project Mentor 2:
Gaya Fior
Technical Assistance, ICTP Scientific Fabrication Laboratory
International Centre for Theoretical Physics (ICTP)
gfior@ictp.it

Abstract

This project develops an AI system that converts architectural photos into tactile formats for visually impaired users. It uses SAM2 for building segmentation and Depth-Anything-V2 for depth estimation. Outputs include 2D vector graphics for laser cutting and 3D printable models for tactile use. The system works with various architectural styles and provides a Streamlit web interface for non-technical cultural heritage staff.

Key results: high building detection accuracy, reliable watertight 3D models, efficient dual-path processing, and fabrication-ready files.

Keywords: Tactile Graphics, Computer Vision, SAM2, Depth Estimation, 3D Reconstruction, Accessibility, Cultural Heritage, Digital Fabrication

Project Objectives and Goals

The goal is to build an AI system that converts photos of buildings and monuments into tactile formats for visually impaired users. This addresses a gap in cultural heritage accessibility using advanced computer vision.

Core Technical Objectives:

Produce high-precision 2D vector graphics (SVG) for laser-cut tactile maps. Generate detailed 3D printable models (STL or OBJ) for tactile exploration and education.

Societal Impact Goals:

The system supports ICTP Fablab and KAUST Academy in making cultural heritage more accessible through low-cost digital fabrication. It lowers technical barriers and reduces production costs for tactile educational materials. This allows museums, cultural institutions, and schools to offer inclusive experiences for visually impaired visitors and students worldwide.

Current Progress and Milestones Achieved

Complete System Development:

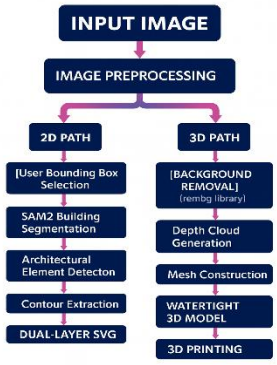
The project has met all main objectives through full system development. SAM2 and Depth-Anything-V2 were integrated into a dual-path architecture for both 2D and 3D outputs. A production-ready Streamlit web interface with advanced user controls was built, and the system was validated on a wide range of architectural images and fabrication setups.

Advanced AI Integration:

SAM2 was used for precise building segmentation and detection of architectural elements across different styles. Depth-Anything-V2 provided high-quality monocular depth maps for 3D generation. Preprocessing pipelines combining multiple vision techniques were developed to enhance model performance.

Production System Deployment:

The final system uses separate pipelines for 2D vector graphics and 3D model generation. The web interface delivers real-time feedback, adjustable parameters, and strong error handling. Quality assurance checks ensure compatibility with multiple types of fabrication equipment.



[Figure 1 - Complete System Architecture Overview]

Methodology and Approach

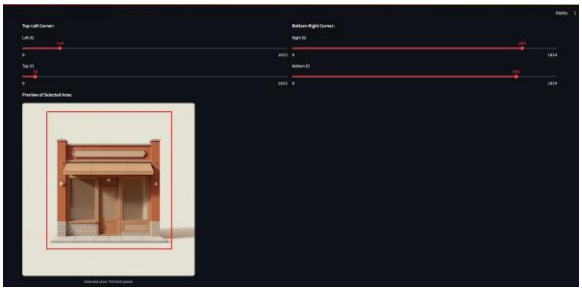
2D Tactile Graphics Generation Pipeline

Step 1: Advanced Image Preprocessing. The 2D pipeline starts with image enhancement using OpenCV. Bilateral filtering reduces noise while preserving edges. CLAHE improves local contrast. Unsharp masking increases detail visibility. These methods improve image quality for AI analysis and keep important architectural features intact.



[Figure 2 - Image Preprocessing Results original vs enhanced]

Step 2: Interactive Building Area Selection. Users select building regions through an interactive web interface with bounding box controls. Sliders allow precise adjustments. This bounding box is key for SAM2 to isolate the building and generate its mask, removing all other elements from the scene.



[Figure 3 - Interactive Bounding Box Selection Interface]

Step 3: Primary Building Segmentation. SAM2 uses the user-defined bounding box to perform primary segmentation. It detects the full building structure and produces an accurate mask. This step isolates the building from the background and surrounding elements, creating a clean base for detailed architectural analysis.



[Figure 4 - Primary Building Segmentation with precise masks]

Step 4: Detailed Architectural Element Detection. A secondary mask generation step detects architectural elements such as windows, doors, columns, and decorative features within the selected building area. Quality filtering with adjustable IoU thresholds and area-based rules refines the results. The user can adjust parameters for more aggressive detection or a lighter pass. This keeps relevant elements and removes noise across different architectural styles and complexity levels.



[Figure 5 - Architectural Element Detection showing detailed features]

Step 5: Vector Conversion. OpenCV contour extraction converts segmentation masks into precise vector paths. Douglas-Peucker polygon simplification, tuned for laser cutting constraints, reduces path complexity. Any open contours are closed to ensure clean shapes for fabrication. The output is a dual-layer SVG separating the full building structure from individual architectural details. This allows flexible fabrication methods such as raised engraving and multi-material production.



[Figure 6 - Final Dual-Layer SVG Output ready for fabrication]

Step 6: Fabricated Tactile 2D Output. The final tactile 2D representation of the building facade is fabricated through laser cutting. This physical model validates the digital workflow, ensuring that architectural details such as windows, doors, and structural outlines are accurately preserved for use in accessibility contexts.



[Figure 7 - Fabricated Tactile 2D Building Output]

3D Model Generation Pipeline

Step 1: Shared Preprocessing Foundation. The 3D pipeline utilizes the same advanced image preprocessing techniques as the 2D pipeline, including bilateral filtering, CLAHE enhancement, and unsharp masking to optimize image quality for depth analysis.

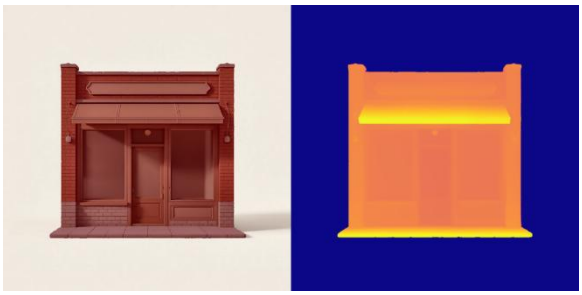
Step 2: Intelligent Background Removal. The rembg library provides automated background removal with architectural-specific optimization parameters, creating clean building isolation that is essential for accurate depth estimation. This process eliminates complex backgrounds and environmental distractions, focusing computational resources on the primary architectural subject.



[Figure 8 - Background Removal Results for 3D Processing]

Step 3: Primary Building Segmentation. SAM2 performs building structure detection using the rembg result that removed background, generating precise masks that define the overall structure boundaries for 3D analysis.

Step 4: Monocular Depth Estimation. Depth-Anything-V2 (ViT-Large architecture) performs state-of-the-art monocular depth estimation on the preprocessed and background-removed images, generating high-fidelity depth maps from single-view architectural imagery. The transformer-based model with 1B+ parameters provides detailed depth information while maintaining architectural feature integrity and spatial relationships.



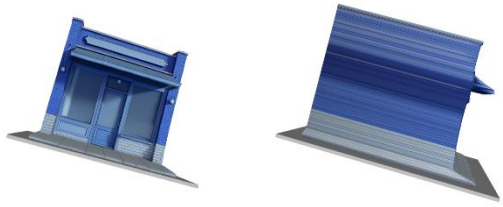
[Figure 9 - High-Quality Depth Map Generation Results]

Step 5: Depth Map Cleaning and Optimization. Sophisticated cleaning algorithms incorporate morphological operations, erosion-based noise removal, and connected component analysis to isolate primary building structures while eliminating background artifacts and depth estimation inconsistencies. This process ensures clean, manufacturable depth information.

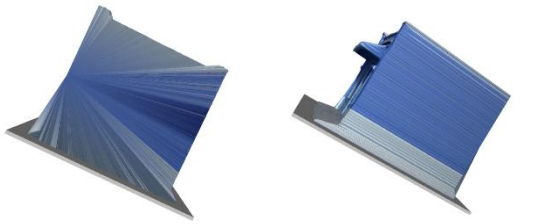
Step 6: 3D Point Cloud Generation. Advanced coordinate transformation algorithms create detailed 3D geometric representations from cleaned depth maps. This process maintains color information from original photographs to ensure visual and tactile coherence in final outputs while optimizing point density for efficient processing.

Step 7: Mesh Construction and Triangulation. Enhanced triangulation algorithms using Open3D libraries convert point cloud data directly to optimized 3D meshes. The process balances triangle density for printing compatibility while preserving essential architectural details and maintaining geometric accuracy.

Step 8: Watertight Model Creation and Quality. Assurance Robust algorithms create closed, manufacturable 3D models through intelligent boundary edge detection, automated wall generation with configurable thickness parameters, and comprehensive base structure integration. Advanced quality assurance includes statistical outlier removal, mesh simplification, and automatic repair of non-manifold edges.

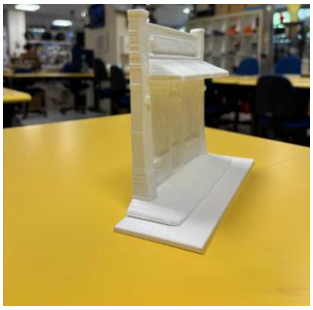


[Figure 10 - Final Watertight 3D Models (Front and right sides)]



[Figure 11 - Final Watertight 3D Models (Back and left sides)]

Step 9: 3D Printing and Tactile Model. The final stage of the pipeline involves exporting the watertight 3D mesh into printable OBJ format. The model is fabricated using a standard 3D printer, ensuring durability and accurate preservation of architectural features. This tactile model enables visually impaired users to physically explore building facades, bridging the gap between digital reconstructions and real-world accessibility.



[Figure 12 - Fabricated 3D Printed Tactile Model]

Web Interface and User Experience. Development The culmination of development efforts resulted in a sophisticated, production-grade web application built using the Streamlit framework. This achievement represents a critical milestone in bridging the gap between complex AI research and practical accessibility applications, making advanced computer vision technology accessible to cultural heritage professionals without requiring technical expertise.

Key Interface Innovations:

- **Intuitive Workflow:** Multi-step process visualization with real-time tracking, clear status updates, and user-friendly feedback.
- **Advanced Controls:** Configurable SAM2 parameters, including grid density, IoU thresholds, and preset modes (Fast/Balanced/Detail) for different use cases.
- **Interactive Selection:** Precise bounding box adjustment with responsive sliders and live previews for accurate region analysis.

Current Implementation Status: The web interface currently supports the complete 2D tactile graphics generation pipeline with full user interaction and real-time processing feedback. The 3D model generation capabilities, while fully functional as standalone processing modules, are not yet integrated into the web interface and require separate execution environments.

Preliminary Results or Findings

2D Pipeline

- **Building Detection:** Excellent results across diverse architectural styles, including classical, modern, and vernacular.
- **Architectural Element Segmentation:** Consistently detected windows, doors, and structural elements.
- **Fabrication Compatibility:** Successfully integrated with laser cutting systems on multiple materials (acrylic, wood, cardboard).
- **Processing Efficiency:** Reasonable runtimes on standard hardware, with effective background noise removal through preprocessing and intelligent segmentation.

3D Pipeline

- **Watertight Meshes:** Reliable generation of closed 3D models suitable for printing.
- **OBJ File Validation:** Verified compatibility with ICTP printing software.

System Usability

- **Efficient User Workflow:** Non-technical users can use the website easily.
- **Parameter Flexibility:** Intelligent preset modes (Fast, Balanced, Detail) optimized results without overwhelming users.

Challenges and Obstacles Encountered

1. Advanced AI Model Integration
    - *Challenge:* Running large-scale model (SAM2) under memory constraints while keeping real-time performance.
    - *Solution:* Implemented caching strategies, checkpoint optimization, and progressive loading. Used torch.inference_mode() to minimize memory usage.
2. 3D Mesh Quality
    - *Challenge:* Early meshes had holes, distortions, and non-manifold edges unsuitable for 3D printing.
    - *Solution:* Built a mesh cleaning pipeline: statistical outlier removal, connected component analysis, watertight reconstruction, and depth-based refinement.

Next Steps and Future Work

- **Multi-View Depth Estimation:** Enhance 3D reconstruction to handle complex geometries and occlusions more accurately.
- **3D Model Generation Integration:** Merge the 3D model reconstruction pipeline directly into the web interface, enabling users to generate both 2D tactile graphics and 3D printable models seamlessly within the same platform.
- **Integration with Museums:** Conduct pilot deployments in educational institutions and cultural heritage organizations to evaluate real-world accessibility impact.

Resources and Tools Used

Deep Learning & AI Frameworks

- **SAM2:** State-of-the-art segmentation model optimized for architectural domains.
- **Depth-Anything V2:** Transformer-based depth estimation with ViT-L backbone.
- **PyTorch Ecosystem:** GPU-accelerated inference, quantization, and batch processing.

Computer Vision & Geometry

- **OpenCV:** Advanced image preprocessing and morphological transformations.
- **Open3D:** Mesh manipulation, optimization, and geometry validation.
- **rembg:** Background removal with architectural optimization.
- **NumPy:** Numerical computing backbone for array and matrix operations.

Web & User Experience

- **Streamlit:** Interactive web app framework with advanced session management.

Conclusion & Impact

This project successfully bridges cutting-edge AI research with practical accessibility applications. It delivers a production-ready system for generating tactile 2D and 3D representations of building facades, validated across fabrication workflows. By integrating SAM2, Depth-Anything V2, and a streamlined web interface, the project empowers cultural heritage professionals and visually impaired users to access and interact with architectural content without technical expertise. The system establishes a replicable methodology for AI-driven accessibility in cultural preservation.

Acknowledgments

We gratefully acknowledge the support of the ICTP Scientific Fabrication Laboratory for technical guidance and access to fabrication equipment, the KAUST Academy for program support, and the open-source AI community for developing the foundational models that made this project possible.

References

[1] Kirillov, A., et al. (2024). *SAM 2: Segment Anything in Images and Videos*. arXiv:2408.00714. Meta AI Research.
[2] Yang, L., et al. (2024). *Depth Anything V2*. arXiv:2406.09414. TikTok Pte. Ltd.
[3] Zhou, Q.Y., Park, J., & Koltun, V. (2018). *Open3D: A Modern Library for 3D Data Processing*. arXiv:1801.09847.
[4] Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal, 25(11), 120–123.
[5] Douglas, D.H., & Peucker, T.K. (1973). *Algorithms for*