A Robust Pipeline for Facade Geometry Estimation Using AprilTags and Digital Imaging

Aravind Srinivasaragavan[1], Danya Liu[1], Kepa Iturralde[1,2] and Christoph Holst[1]

[1]Chair of Engineering Geodesy, School of Engineering and Design,
Technical University of Munich, 80333 Munich, Germany
[2]Chair of Digital Transformation in Construction, Institute of Construction Management,
Faculty of Civil and Environmental Engineering, University of Stuttgart, 70569 Stuttgart, Germany

aravind.srinivasaragavan@tum.de, danya.liu@tum.de, kepa.iturralde@ibl.uni-stuttgart.de, christoph.holst@tum.de

Abstract -

The accurate modeling of building facades plays a critical role in architecture, urban planning, and energy efficiency. Traditional methods of facade mapping are often labor-intensive, costly, and prone to error. Recent advancements in computer vision and sensor technology have introduced innovative approaches for capturing precise geometric data, including the use of AprilTags—fiducial markers that enable efficient facade reconstruction through digital imaging techniques. This paper explores a method that involves arranging AprilTags in a grid pattern on building surfaces, capturing multiple images of the tagged facade, and using triangulation techniques to infer the 3D geometry of the structure with high accuracy. The resulting digital models have a wide range of applications, including optimizing solar panel placement, creating detailed replicas for heritage conservation, and enhancing urban planning through improved simulations of light, shadows, and heat.

Keywords -

Facade Modeling; AprilTags; 3D Reconstruction; Stereo Vision; Camera Calibration; Fiducial Markers; Distance Estimation

1 Introduction

The accurate modeling of building facades has become increasingly important in various fields such as architecture, urban planning, and energy efficiency. Traditional methods of mapping and measuring facades can be time-consuming, costly, and prone to error. Recent advancements in computer vision and sensor technology have introduced new approaches for capturing detailed and precise geometric data of structures. One such method involves the use of AprilTags—fiducial [1] markers that can be placed on the facade—and digital imaging techniques [2] to estimate the real-world distances between these markers, effectively reconstructing the facade's geometry.

In this technique, AprilTags are arranged in a grid pattern on the building's surface, either manually or with the assistance of a drone. A camera then captures multiple images of the tagged facade from different angles, and, combined with information such as the camera's focal length and intrinsic parameters, these images are used to infer the relative positions of the tags. By triangulating the distances between the tags, the method can reconstruct the 3D geometry of the facade with high accuracy. This process allows for the creation of a digital model that closely represents the real-world structure, enabling applications in design, planning, and analysis.

The resulting geometric data has diverse applications beyond simple structural visualization. For example, in the field of renewable energy, the technique can be used to precisely model building surfaces for the optimal placement of solar panels. Similarly, it can assist in heritage conservation, where detailed digital replicas of historical facades can be created for restoration or archival purposes. Furthermore, this method could be applied to urban studies, enabling more accurate simulations of shadows, light exposure, and heat absorption in city planning. As building facades continue to evolve in complexity and functionality, the ability to quickly and accurately model these surfaces opens up new possibilities for both practical and theoretical applications in the built environment.

2 Problem Description and Constraints

The accurate modeling of building facades is a critical task in fields such as architecture, urban planning, and energy efficiency. However, traditional methods of mapping and measuring building facades are inherently challenging due to several factors. These methods are often time-consuming, labor-intensive, and prone to human error, especially when dealing with large or complex structures. Furthermore, the equipment required for manual measurements can be costly and may not always yield the level of precision necessary for advanced applications.

Recent advancements in computer vision and sensor technology offer an opportunity to address these challenges [3]. Specifically, the use of AprilTags—fiducial markers designed for robust detection—has emerged as a promising approach for facade modeling. These markers enable precise distance measurements and allow for the reconstruction of a facade's geometry through digital imaging.

First, the placement of AprilTags requires a systematic and controlled process, typically arranged in a grid pattern on the building's surface. This can be done manually or with the assistance of drones, but both approaches present logistical challenges, particularly for large buildings or difficult-to-access facades. Second, the accuracy of the reconstruction depends heavily on the quality and quantity of the images captured. Factors such as camera calibration, lighting conditions, and the angles from which the images are taken can significantly impact the precision of the results. Additionally, the computational process of triangulating the positions of the tags requires careful handling of camera intrinsic parameters, such as focal length, to ensure reliable 3D geometry generation.

These constraints highlight the need for a method that balances practicality and accuracy while minimizing costs and manual effort. Despite these challenges, the use of AprilTags for facade modeling represents a scalable and efficient solution that could potentially overcome the limitations of traditional techniques, enabling diverse applications in design, analysis, and planning. In short, the main problem is to determine the spacial geometry of the placed AprilTags as accurately as possible. The spacial geometry of the Apriltag grid directly gives one information about the geometry of the building facade. In this paper, we propose a pipeline, which only involves a Camera with a fixed focal length and a Python pipeline, which can be used to very accurately determine the distance to an Apriltag. This information can then be used to estimate the spacial geometry of the AprilTag grid.

3 Technical Preleminaries

In the process of reconstructing the geometry of building facades, one common challenge is the presence of noise and inaccuracies in the observed 3D data points, especially when these points are assumed to be co-planar. While techniques like AprilTag-based distance estimation offer high accuracy in reconstructing building geometry, small errors in the measurements or in the camera calibration can introduce deviations that affect the precision of the reconstructed model. In the next section, we will discuss one possible method to mitigate this issue.

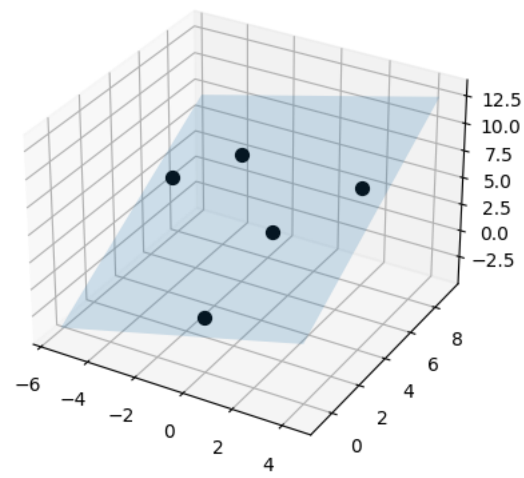

Figure 1. The Best-Fitting plane passing through 5 points

3.1 Best-Fitting Plane: The Key Improvement Strategy

A practical solution to this problem is the concept of the best-fitting plane—a mathematical approach that minimizes the error between observed data points and a plane that best represents their true position.

The best-fitting plane is typically computed using methods like least squares fitting, where the goal is to find the plane that minimizes the sum of squared distances between the observed data points and the plane itself. This method is particularly useful when dealing with 3D data points that are known a-priori to lie approximately on the same plane, such as those derived from the AprilTag markers on a building facade. By fitting a plane to these points, we can correct for small inaccuracies in their positioning, leading to a more accurate representation of the facade geometry.

In the context of our research, using a best-fitting plane can significantly enhance the quality of the 3D model by refining the locations of the AprilTags as shown in Figure 1. This correction process helps mitigate errors caused by slight misalignments or measurement noise, ensuring that the reconstructed facade closely matches the actual structure. Such improvements are particularly valuable when the facade's geometry is needed for precise applications like solar panel placement, where even minor discrepancies in the 3D model could affect performance. By leveraging the best-fitting plane approach, this method can contribute to more reliable and accurate models for various real-world applications, including energy optimization, urban planning, and architectural analysis.

3.2 Distance Estimation via Effective Focal Length

Accurate distance estimation to an AprilTag from a camera is a critical aspect of the proposed facade modeling method. This process relies on key camera parameters, including the *effective focal length* and the *pixel size* of the camera's sensor. These parameters enable the calculation of real-world distances by relating the apparent size of the AprilTag in the captured image to its known physical dimensions.

3.3 Effective Focal Length and Camera Geometry

The effective focal length (f) of a camera [4] is the distance from the lens to the sensor at which parallel rays of light converge to form a sharp image. It is typically expressed in millimeters and is a fundamental property of the camera's optics. Combined with the sensor's pixel size (p), the focal length determines the *field of view (FOV)* and directly impacts the perceived size of objects in the image.

For a camera with a focal length f and a sensor with pixel size p, the physical distance between the camera and an object, such as an AprilTag, can be determined using the principle of similar triangles. The relationship between the tag's real-world dimensions and its image size on the sensor provides the basis for this calculation.

3.4 Calculating Distance Using Image Data

When an AprilTag is captured in an image, its size in pixels (s_{pixels}) can be measured. The real-world width of the tag (W) is known a priori. Using these values, the distance to the tag (D) can be calculated using the following formula:

$$D = \frac{f \cdot W}{s_{\text{pixels}} \cdot p}$$

Here:

- f is the effective focal length of the camera (in millimeters),
- W is the real-world width of the AprilTag (in millimeters),
- s_{pixels} is the width of the AprilTag in the image (in pixels),
- p is the physical size of a single pixel on the sensor (in millimeters).

3.5 Practical Considerations

1. **Camera Calibration:** Accurate distance estimation requires precise knowledge of the camera's intrinsic parameters, including the focal length and pixel size. Camera calibration processes, such as those using checkerboard patterns, are essential for determining these values accurately.
2. **Tag Orientation and Perspective:** If the AprilTag is not perpendicular to the camera's optical axis, its apparent size in the image will be distorted. Perspective correction algorithms may be required to account for this distortion and ensure accurate measurements.
3. **Resolution and Image Quality:** The resolution of the captured image affects the precision of s_{pixels}. High-resolution images provide more accurate measurements of the tag's size, improving distance estimation accuracy.
4. **Lighting and Focus:** The clarity and visibility of the AprilTag in the image depend on proper lighting and focus. Poor lighting conditions or out-of-focus images can introduce errors in size estimation and, consequently, distance calculations.

3.6 Applications in Facade Modeling

By determining the distances to multiple AprilTags placed on a building's facade, a 3D reconstruction of the surface can be performed. The calculated distances form the basis for triangulating the relative positions of the tags, enabling the generation of a detailed geometric model of the structure. This approach, leveraging the effective focal length and pixel size, provides a robust and scalable solution for facade modeling with high accuracy.

4 The Main Pipeline

The proposed pipeline for facade modeling leverages AprilTags and digital imaging techniques to accurately reconstruct the geometry of building facades. The steps involved in the pipeline are outlined below:

4.1 Placement of AprilTags

Several AprilTags are placed in a grid-like formation on the facade like Figure 2.

This systematic arrangement ensures that the tags are evenly distributed across the surface, enabling accurate reconstruction. Care is taken to maintain consistent spacing between the tags and to cover the entire area of interest.

4.2 Image Capture

Multiple images of the tagged facade are captured from different distances. To minimize distortion and maintain accuracy, the camera angles are kept relatively shallow, avoiding overly steep perspectives that could negatively affect the results. These images serve as the input for the subsequent processing and analysis stages. For the

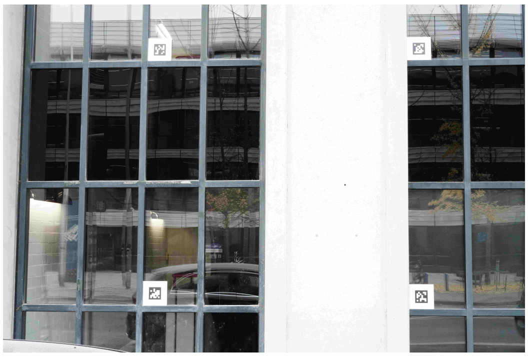

Figure 2. The setting of the AprilTags' outline

Figure 3. Contrast Adjusted Image

experiments themselves, we used the Sony Alpha 7R IV 35-mm-Vollbildkamera with 61,0 MP [5]. Based on our experiments, we found that this particular camera has an effective focal length of approximately 13257.6 (in the appropriate units).

4.3 Pre-Processing

Before detecting the AprilTags, the captured images undergo pre-processing to enhance their quality. This step includes modifying parameters such as gamma and contrast to ensure the tags are clearly visible and detectable in the images as shown in Figure 3.

These adjustments help mitigate the effects of varying lighting conditions and improve the reliability of the detection process.

4.4 AprilTag Detection

The AprilTags in the pre-processed images are detected shown in Figure 4 using the Python library `AprilTag` [6], which is available at `https://github.com/Tinker-Twins/AprilTag`. This library identifies the tags and determines their pixel dimensions (s_{pixels}) in the images. These pixel measurements are essential for calculating the real-world distances to the tags.

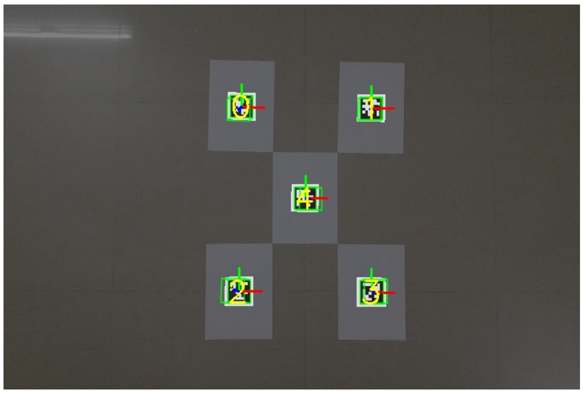

Figure 4. Detection results

4.5 Calculating the Distance Between Two AprilTags

Once the distances from the camera to individual AprilTags and their positions in the 3D space are determined, the distance between two AprilTags can be calculated by analyzing their relative spatial coordinates.

4.6 Steps for Calculation

1. **Determine the 3D Coordinates of Each AprilTag:** Using the camera as the origin of the coordinate system, the 3D position of each AprilTag (x, y, z) can be determined from the camera parameters and the pixel coordinates of the tag in the image. For a tag at distance D from the camera:

$$x = D \cdot \frac{u - u_0}{f_x}, \quad y = D \cdot \frac{v - v_0}{f_y}, \quad z = D$$

   where:

   - D is the distance from the camera to the AprilTag (calculated using the focal length, tag size, and pixel size),
   - (u, v) are the pixel coordinates of the AprilTag in the image,
   - (u_0, v_0) are the principal point coordinates (typically the center of the image),
   - f_x and f_y are the focal lengths of the camera in the x- and y-axes (in pixel units).

2. **Final Correction via the Best-Fitting Plane** Since we know (approximately), that the AprilTags all lie in a Plane, we can use the best fitting plane for the 3D coordinates obtained in the previous section and project the points on this plane to obtain the corrected points (x', y', z'). Note that we used the Python library `scikit-spacial` [7] for the Best-Fitting plane compuations.

3. **Calculate the Euclidean Distance Between Two AprilTags:** Let the coordinates of the first AprilTag be (x_1, y_1, z_1) and the second AprilTag be (x_2, y_2, z_2). The distance d between the two tags is given by the Euclidean distance formula:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

4.7 Practical Considerations

- **Accuracy of Coordinates:** The accuracy of the calculated coordinates depends on the precision of the distance measurements (D) and the camera's intrinsic parameters (f_x, f_y, u_0, v_0).

- **Perspective Distortion:** If the AprilTags are not aligned perpendicularly to the camera's optical axis, corrections may be required to ensure accurate coordinate estimation.

- **Tag Identification:** Each tag must be uniquely identified in the image to correctly associate its 3D position with its real-world counterpart.

By calculating the 3D positions of multiple AprilTags and using the Euclidean distance formula, the relative distances between tags can be computed with high precision. This process enables the reconstruction of the facade's geometry with reliable spatial relationships between the tags.

5 Experimental Results

Several experiments were performed on one of the TUM buildings in Arcisstrasse by considering a section of the windows as the Facade. Six AprilTags, with different tag IDs were placed on the windows as shown in Figure. In this particular experiment, AprilTags with the following IDs were used: $1, 2, 3, 1617, 18, 20, 22$. Here we describe the results of once such experiment. The results of this experiment are described in two tables. Table 1 shows the estimated distance between the tags Tag A and Tag B (A and B being the Tag IDs) for three specific cases.

On the other hand, Tables 2 and 3 show the the 3D coordinates of the Tags with ID 2 and 16. Here, the Tag 1 is the reference Tag. This means that the coordinates have been changed in such a way so that the Tag 1 becomes the origin (with coordinates $(0, 0, 0)$).

Distance Between Tags(mm)			
Tag A and Tag B	3-18	16-22	20 - 22
1	1235.3	2106.9	1621.0
2	1248.7	2127.7	1653.5
3	1278.7	2210.4	1605.1
4	1270.9	2183.5	1734.3
Ground Truth	1250.0	2150.0	1650.0

Table 1. Estimated distance between the tags

Tag 2 Coordinates (mm)			
Axis	X	Y	Z
1	-4979.87	61.68	63.04
2	-4979.50	53.34	55.96
3	-3345.30	71.55	86.37

Table 2. Estimated Coordinates of a Tag

6 Conclusion and Future Developments

In this work, we presented a robust pipeline for the accurate modeling of building facades using AprilTags and digital imaging techniques. By leveraging fiducial markers arranged in a systematic grid pattern, we demonstrated how the real-world distances and relative positions of these tags can be calculated to reconstruct the 3D geometry of facades. The process incorporates critical steps such as camera calibration, pre-processing of images to enhance quality, and the use of the `AprilTag` Python library for precise tag detection. The geometric data obtained from this pipeline can be applied to a wide range of fields, including renewable energy optimization, heritage conservation, and urban planning. One more thing to note is that this pipeline does not involve complicated (and rather archaic) tools such as a Total Station. Further, this pipeline can reduce the time it takes to estimate the facade geometry in comparison to other traditional methods.

We also discussed the mathematical foundations underpinning the reconstruction process, particularly the use of the camera's effective focal length, pixel size, and image data to determine the spatial coordinates of each AprilTag. Furthermore, the ability to calculate the distances between tags using their 3D positions allows for the creation of highly accurate and detailed facade models.

Looking ahead, we plan to enhance the accuracy and scalability of our approach by incorporating stereo vision techniques [8, 9, 10, 11]. Specifically, future experiments will utilize the FUJIFILM GFX 100 II camera [12], a high-resolution imaging system, in a dual-camera setup(the other camera being the current on). By capturing synchronized images from two cameras at different perspectives, stereo vision will enable precise depth estimation and reduce potential errors caused by distortion or

Tag 16 Coordinates (mm)			
Axis	X	Y	Z
1	135.63	3812.57	63.04
2	127.26	3808.09	55.96
3	136.23	3725.82	86.37

Table 3. Estimated Coordinates of a Tag

steep angles. This will significantly improve the reliability of 3D reconstructions and make the method more suitable for complex facade geometries.

The integration of advanced camera systems and stereo vision techniques represents a promising direction for this work, offering increased accuracy and flexibility for facade modeling. These advancements will further expand the applicability of our approach to a broader range of architectural, engineering, and urban development challenges.

References

[1] Yifan Zhang, Jian Wang, and Xinyu Li. Analysis and improvements in apriltag based state estimation. *Sensors*, 23(1):123, 2023.

[2] Zhen Li, Jian Zhang, Ming Li, and Jun Wang. Automatic reconstruction of building façade model from photogrammetric mesh model. *Remote Sensing*, 13 (19):3801, 2021.

[3] Kepa Iturralde, Samanti Das, Aravind Srinivasaragavan, Thomas Bock, and Christoph Holst. An automated prefabricated facade layout definition for residential building renovation. *Buildings*, 13(12), 2023. ISSN 2075-5309. URL `https://www.mdpi.com/2075-5309/13/12/2981`.

[4] Canon Europe. *Understanding Focal Length*, 2025. URL `https://www.canon.ge/pro/infobank/understanding-focal-length/`. Accessed: 2025-01-15.

[5] Sony Europe B.V. Alpha 7R IV 35-mm-Vollformatkamera mit 61,0 MP description, 1999. URL `https://www.sony.de/electronics/wechselobjektivkameras/ilce-7rm4`.

[6] Tinker-Twins. Apriltag python library. `https://github.com/Tinker-Twins/AprilTag`, 2025.

[7] Andrew Hynes. *scikit-spatial: Spatial objects and computations based on NumPy arrays*, 2023. URL `https://scikit-spatial.readthedocs.io/en/stable/api_reference/Plane/methods/skspatial.objects.Plane.best_fit.html`. Accessed: 2025-01-15.

[8] Liang Tang and Gerard Medioni. Building facade detection, segmentation, and parameter estimation for mobile robot stereo vision applications. *Image and Vision Computing*, 31(10):759–773, 2013.

[9] Graham Fyffe, Andrew Jones, Oleg Alexander, and Paul Debevec. Multi-view stereo on consistent face topology. *ACM Transactions on Graphics (TOG)*, 36 (4):1–10, 2017.

[10] Hao Wang, Wei Li, and Yiming Zhang. Landmark-based localization using stereo vision and deep learning in gps-denied environments. *arXiv preprint arXiv:2402.12551*, 2023.

[11] Jeffrey A Delmerico, Jason J Corso, and Philip David. Boosting with stereo features for building facade detection on mobile platforms. In *Proceedings of Western New York Image Processing Workshop*, pages 1–4. IEEE, 2010.

[12] FUJIFILM Corporation. *FUJIFILM GFX100 II Product Page*, 2025. URL `https://eshop.fujifilm-x.com/de/fujifilm-gfx100-ii.html`. Accessed: 2025-01-15.

Acknowledgement

The research presented in this paper was originally carried out in the context of the H2020 project **ENSNARE**. **ENSNARE** has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. **958445**.