MPI Informatics Building Model as Data for Your Research

Vlastimil Havran

Czech Technical University in Prague
Department of Computer Graphics and Interaction (DCGI)
http://dcgi.felk.cvut.cz/~havran

joint work with

Jozef Zajac (MPII, Slovakia)
Jiri Drahokoupil (DCGI)
Hans-Peter Seidel (MPII)

DCGI

max planck institut informatik

Talk Outline

- Video for animated model
- Take home message
- Project motivation
- Related projects
- Input and output data
- Model accuracy and level of detail
- Exploitation and algorithmic problems
- Conclusion and future work

Take Home Message

- The 3D data of MPI Informatics building including luminaires, hand estimated surface reflectances, and reference HDR images are available on the web:

http://www.mpi-inf.mpg.de/resources/mpimodel/v1.0/

- The data complexity is high (over 73 millions triangles, 623 light sources, and 53 surface reflectances)
- Several available 3D data formats (3D Studio MAX, VRML 97, extended VRML97, PBRT)
- Data can be used for non-commercial research for free without applying for any licence, permission etc.
- There is a technical report and presentation slides describing the project flow, problems, results, and additional data

Project Motivation

- Data are needed for algorithms as they are subject of algorithms they process them. Some research topics are solely dependent on the data.
- There were some 3D data sets in the previous work, but there are always some issues to consider…

Image courtesy of Wald et al., 2004

Algorithm Performance Dependence on Data

- Size (=number of elements), where element has to be defined in advanced - real numbers sorting in O(N log N)
- Data dimensionality
- Some other known data properties
  - limited precision or representation - radix sort in O(N)
  - statistical properties – partially ordered sequences O(?)
  - information content, entropy,…
  - others ….

- General problems with data
  - incompleteness for some purposes.
  - open data format: powerful, flexible, and simple?

Note: serious ethic problems using private company data (tracking authors of submitted papers, real reproducibility)

Reproducibility of Research Work

- Research work has to be reproducible

- **Wikipedia:** “Reproducibility is one of the main principles of the scientific method, and refers to the ability of a test or experiment to be accurately reproduced, or replicated, by someone else working independently.”

- **Rules of Good Scientific Practice, Max Planck Society, 24. November 2000:**

  “The primary test of a scientific result is its reproducibility. The more surprising, but also the more desirable a result is, the more important it is – as far as is possible with justifiable expense or effort – that the route to that result be independently repeated within the research group before the result are passed on to the outside.”

  (Cited text at: http://www.mpg.de/pdf/rulesScientificPract.pdf)

#1 Some Datasets already Small or Obsolete

- Examples:

Utah teapot
Year 1975
~ 10,000 triangles, originally modeled by Bezier surfaces

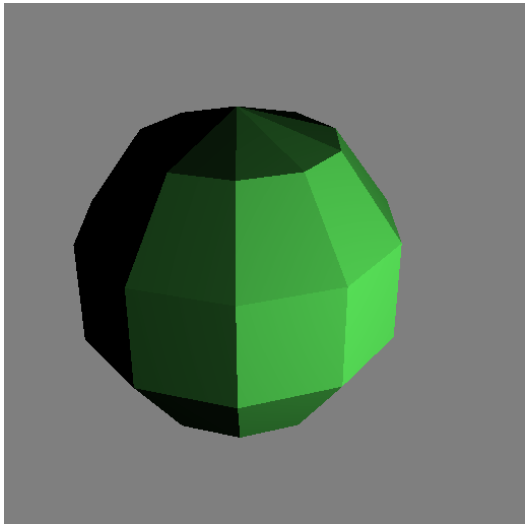
? Stanford bunny ?
Year 1993
69,451 triangles

The Stanford 3D Scanning Repository

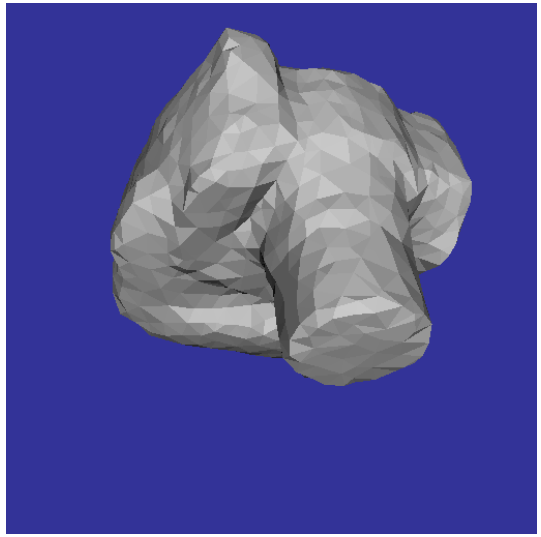
- Nice models anyway, but not really competitive with the real data complexity used in practice in 2010.
- Hypothesis: Each dataset will become obsolete in future (including our project).

#2 - Low Quality/Unrealistic/Uninteresting/Unconvincing?

73 primitives

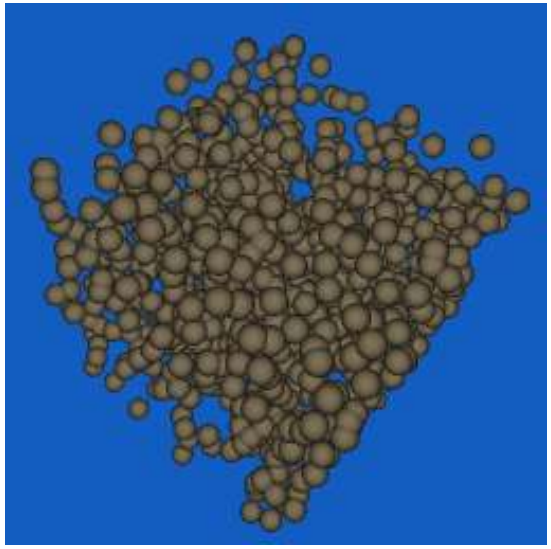


2,324 primitives

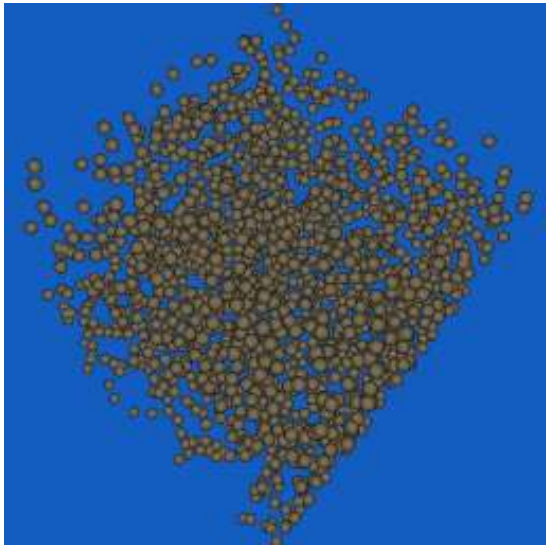


Poisson point process

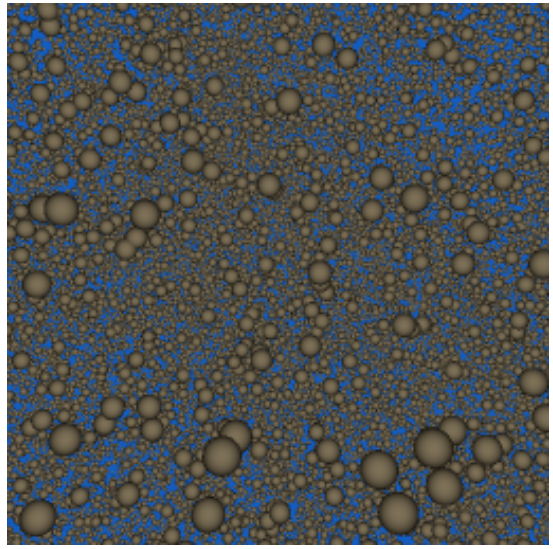
10000 primitives



2,000 primitives

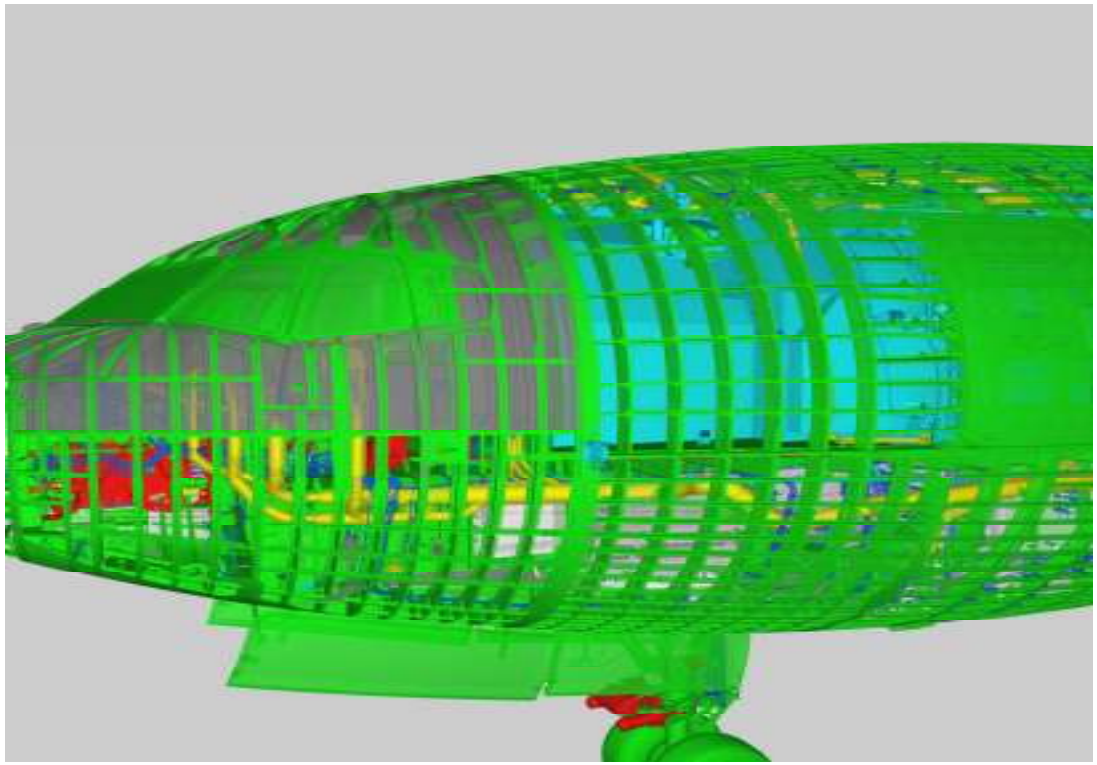


100,000 primitives



#3 - Nice 3D datasets - Rarely Available from a Company

Rare Exception: Boeing model, 350 millions of triangles



However, incomplete model for serious use in **predictive image synthesis:**

- no reflectance data
- no luminaire data
- no acquired reference photographs for controlled outdoor lighting conditions, a difficult access to the airplane.

Some Known Available 3D Data Sets

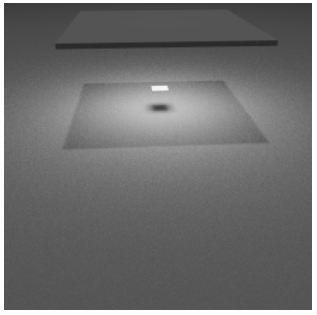
- Stanford 3D scanning repository – individual objects, 1993
  http://graphics.stanford.edu/data/3Dscanrep/
- Soda Hall building model by Prof. Séquin, 1993
  http://www.eecs.berkeley.edu/~sequin/soda/soda.html (558,353 surfaces)
- BES project repository for ray shooting, 1999
  http://www.cgg.cvut.cz/members/havran/BES/
  – Trial to collect useful 3D datasets on the web
- Sibenik cathedral model and Sponza Atrium model, Marko Dabrovic, 2002, http://hdri.cgtechniques.com/~sibenik2/
- AIMshape object repository, 2005
  http://shapes.aimatshape.net/

Validation Methods for Global Illumination Computation Related Work and Datasets before Year 2003

- *Testing Monte-Carlo Global Illumination Methods with Analytically Computable Scenes*, 2001, Prof. Szirmay-Kalos, Budapest.
- Cornell Box Project, 1998

  http://www.graphics.cornell.edu/online/box/
- Drago and Myszkowski's Atrium Project, 1999 (next slide)
- *Global Illumination Test Scenes,* Smits and Wann-Jensen, 2000, technical report

  http://www2.cs.utah.edu/~bes/papers/scenes/



Complex Architectural Dataset for Predictive Rendering

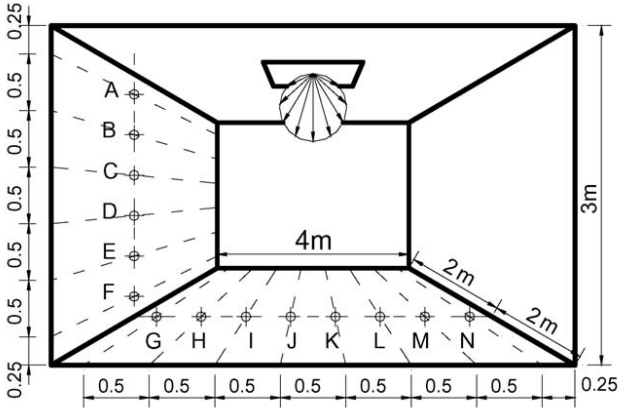
- Aizu Atrium, project by F. Drago and K. Myszkowski, 1999, hosted at MPII http://www.mpi-inf.mpg.de/resources/atrium/

  Published as “Validation Proposal for Global Illumination and Rendering Techniques”, journal C&G, 2001

- Subtle or serious problems with AIZU Atrium Model?
  - Is a problem for EU IST project that the building is in Japan?
  - Could be spatial reflectance data (BTF, 7D function) measured additionally?
  - Data were not created in the scope of RR project, is this a problem?
  - Sufficiently complex? Data only for a single hall, no data behind the doors.

Validation Methods for Global Illumination Computation Related Work and Datasets before Year 2003

- Roland Schregle PhD Thesis, 2005, Daylight Simulation with Photon Maps
- Ulbricht, Wilkie, Purgathofer: Verification of Physically Based Rendering Algorithms, EG STAR 2005
- Havran, Myszkowski, Wurster, Soler: D10.3 Report on Light Simulation in Architecture, 2005, RR project
- CIE 171:2006 Test Cases to Assess the Accuracy of Lighting Computer Programs



Real Reflect EU IST Project – Let us use MPI Informatics Building?

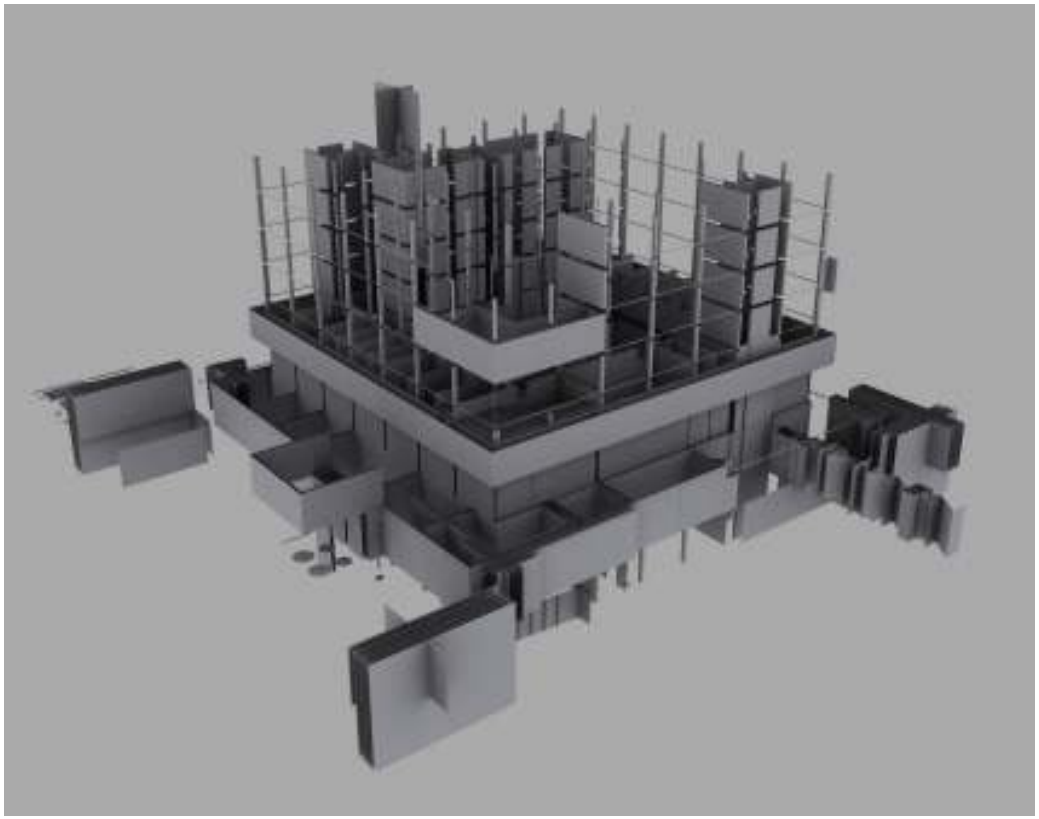
- February 2003 - let us try to use AIZU Atrium and also virtually reconstruct MPII building, it should be possible before the end of RR project or earlier, DXF model should exist!
- Some images / video were known to be rendered in the past by architects of the MPII building:

Photograph

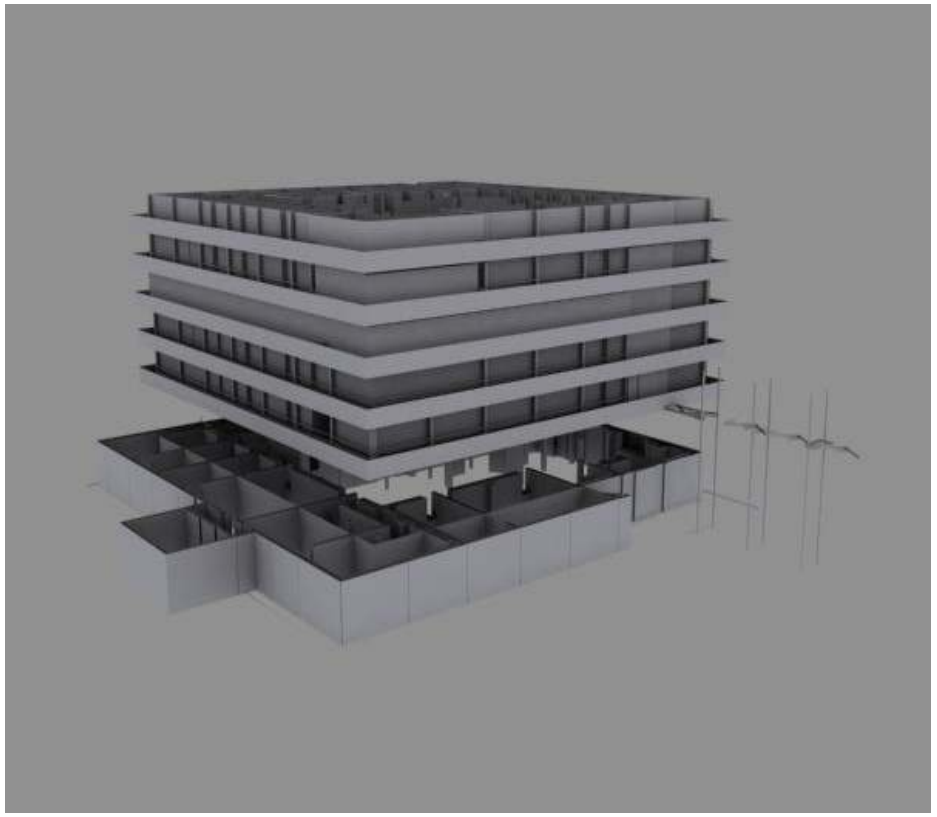
Rendered image, year ~1993

Project Input Data - Geometry

- Two assumptions were wrong:
  - *"3D data in DXF files will be obtained quickly."* - 6 months
  - *"3D data from architects will be useful."*
- Software problem – **data migration** from an older software

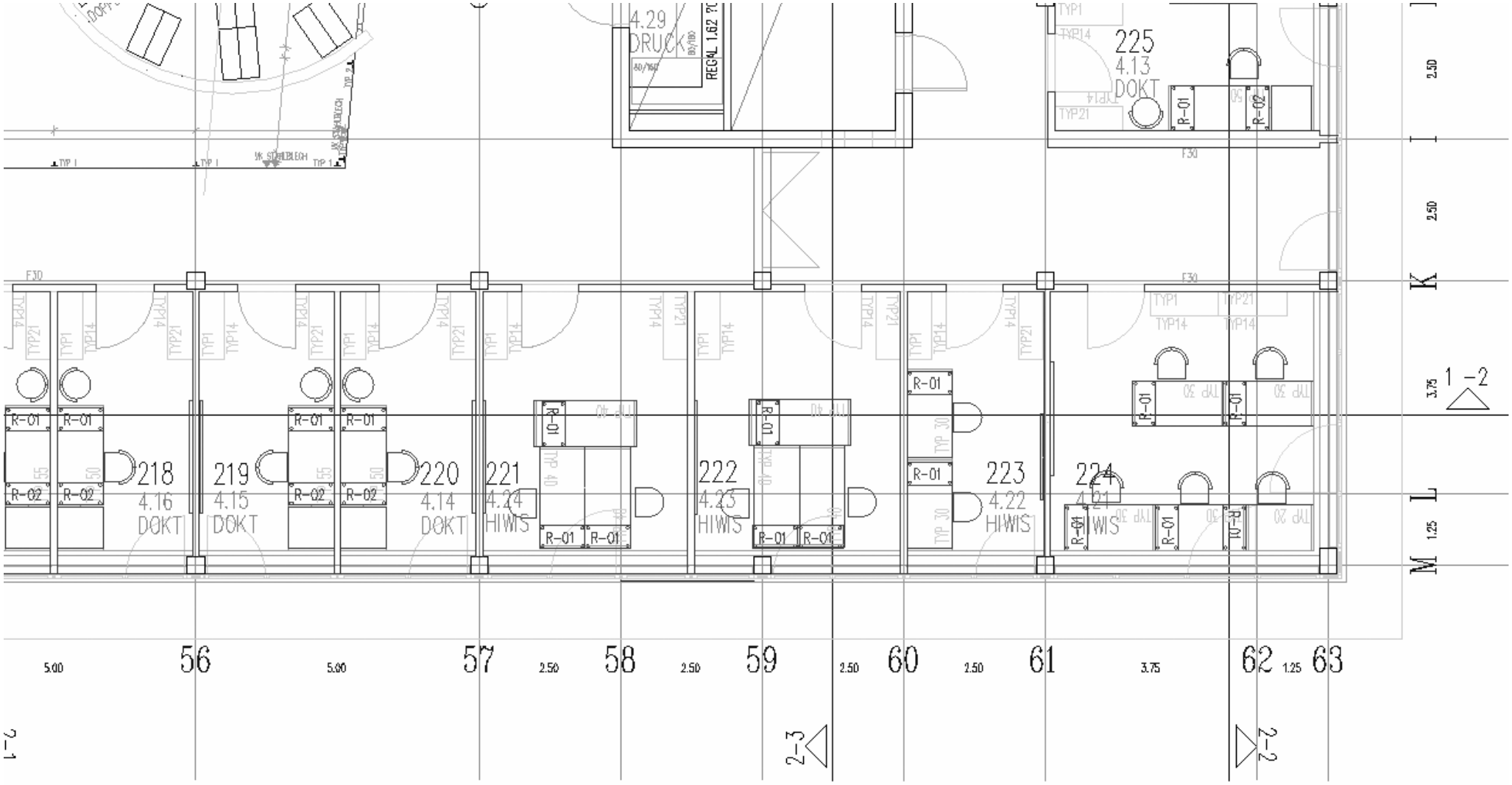


After the data import



After manual repositioning

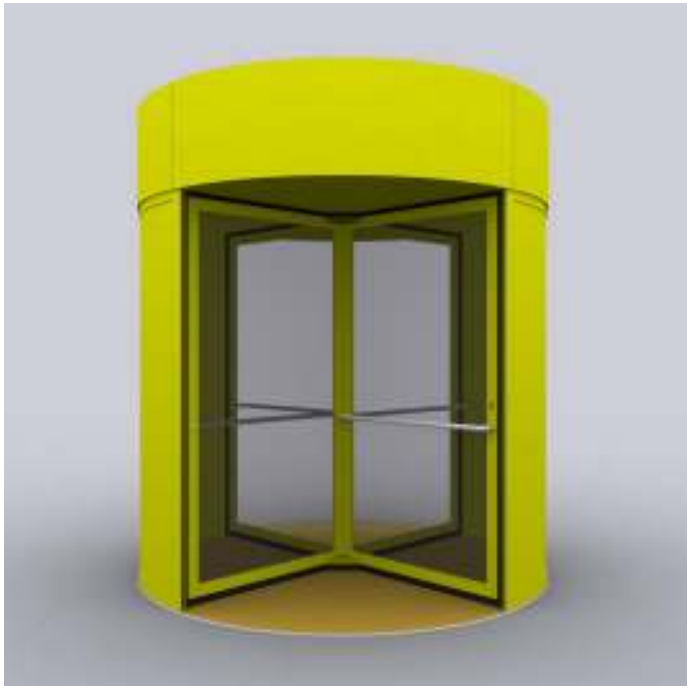
Technical Drawing Example – 2nd floor

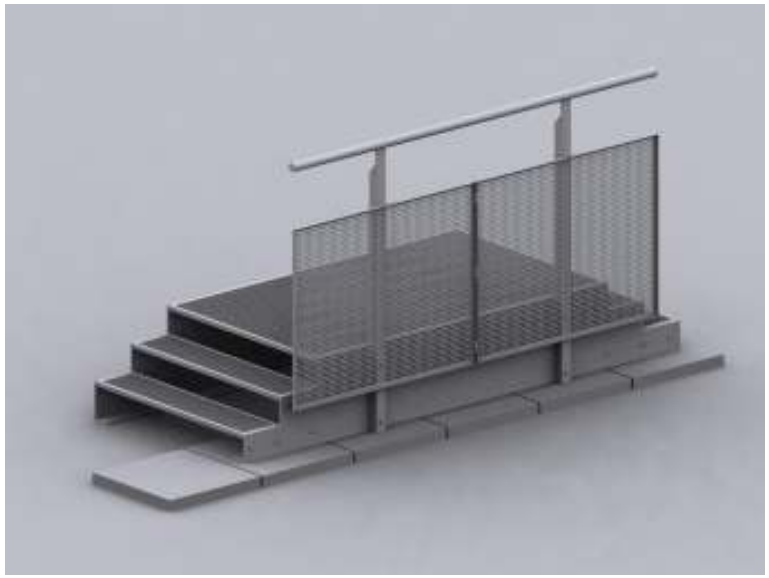


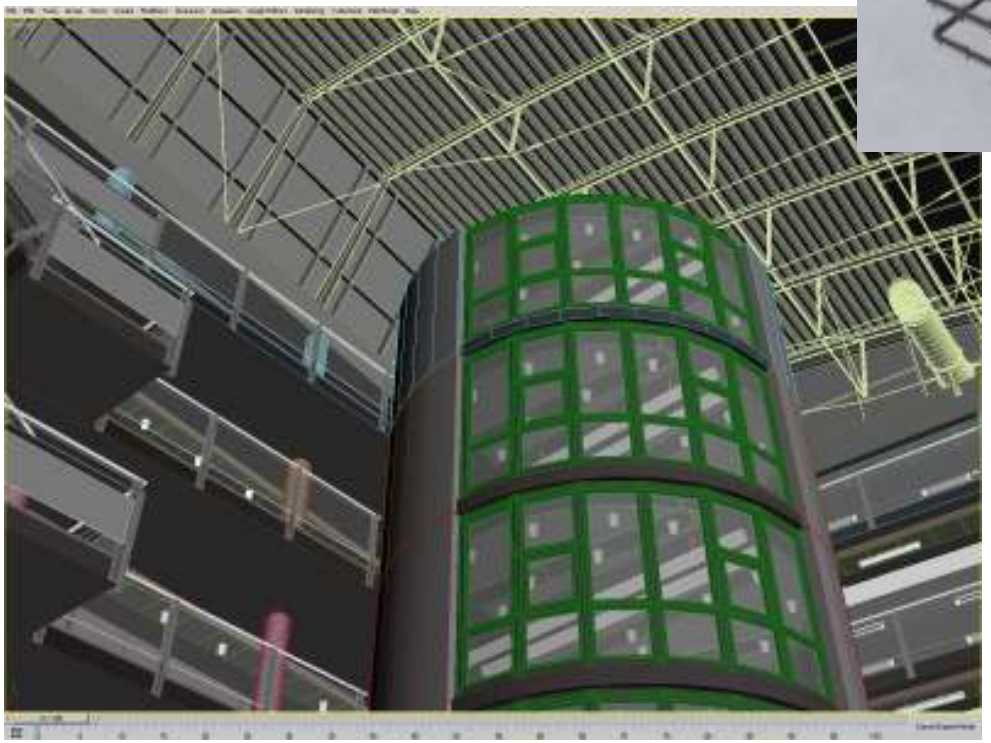
DCGI

max planck institut informatik

Example Results in November 2004

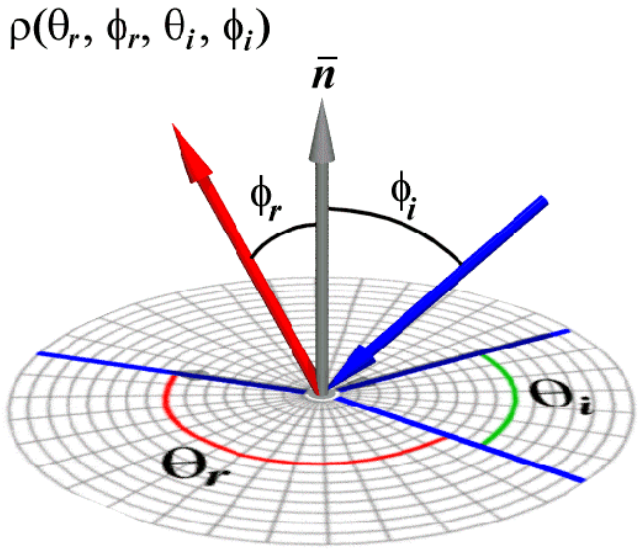






Reflectance (BRDF) Acquisition

- A sample is needed for BRDF measurement, 30x30mm for each material in the building construction …..?!?
- BRDF (bidirectional reflectance distribution function) is a 5-dimensional function, uneasy to measure, represent and use, some efficient data representation is needed.



Gantry for BRDF acquisition
Lightec company (France)

BRDF estimation

- No way to measure the reflectance data with high accuracy
  - Extraction of 63 material samples in the MPII building ?!?
  - Problem to find the same samples without the destruction of the building.
  - Problem to find a cheap way to measure the reflectance.
- Approximate visual appearance for materials known from the photographs was known.
- BRDF estimated from several other reflectance databases or data found on the web, in particular using the project:

  CUReT:Columbia-Utrecht Reflectance and Texture Database,

  http://www1.cs.columbia.edu/CAVE/software/curet (Year 1999),

  Oren-Nayar model for BRDF (used in PBRT software)

Luminaire Data Acquisition

- Directional luminous intensity only 2-dimensional function – easier.
- Some data could be available on the web, however,
  in this project the data were not available, should be measured.
- Manual measurements were found inaccurate and extremely tedious.

Manual luminaire acquisition test, September 2005

goniophotometer Lightec company (France)

Luminaire Measurements for Six Types of Luminaires at Czech Technical University in Prague

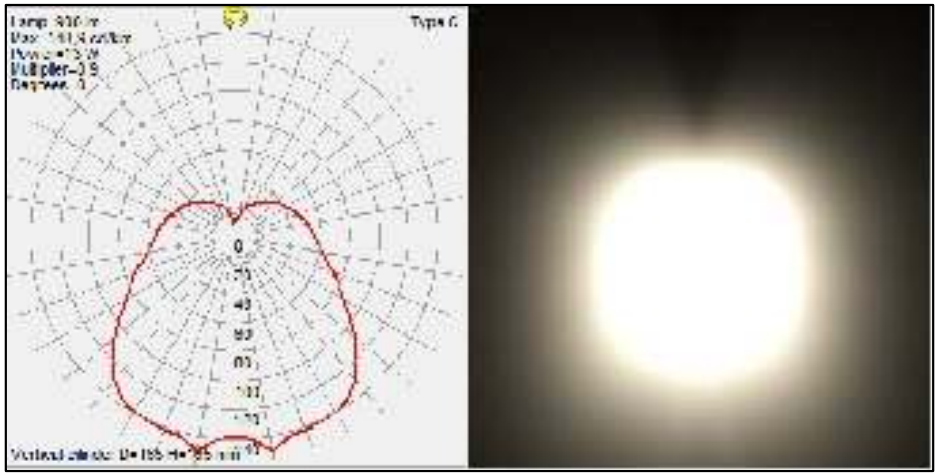
- Help of the department working also in lighting (Prof. Jiri Habel and Ing. Marek Balsky)

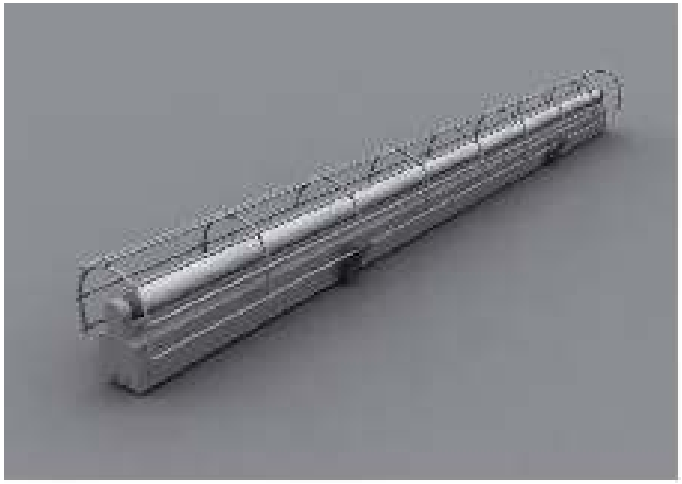
Goniophotometer and light bench used

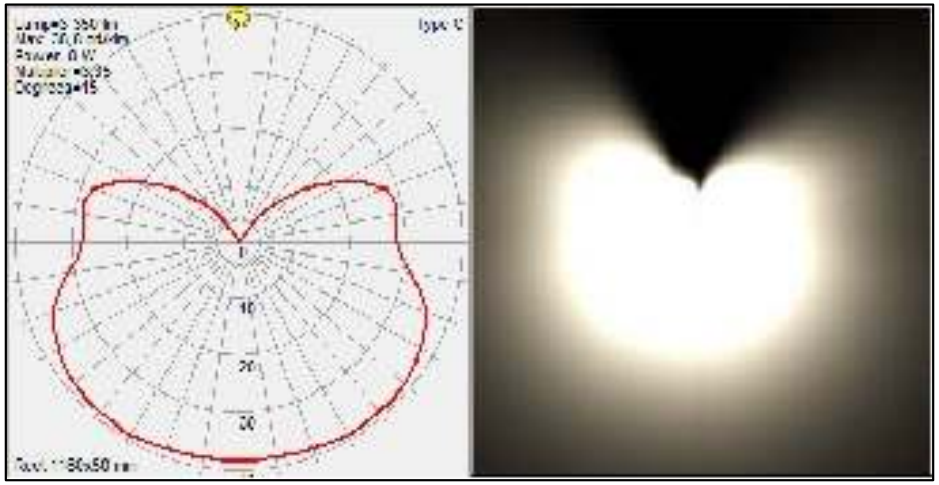
Example: Emission Distribution Data for Two Luminaires

Luminaire geometry

Emission goniometric diagram
(far field photometry)

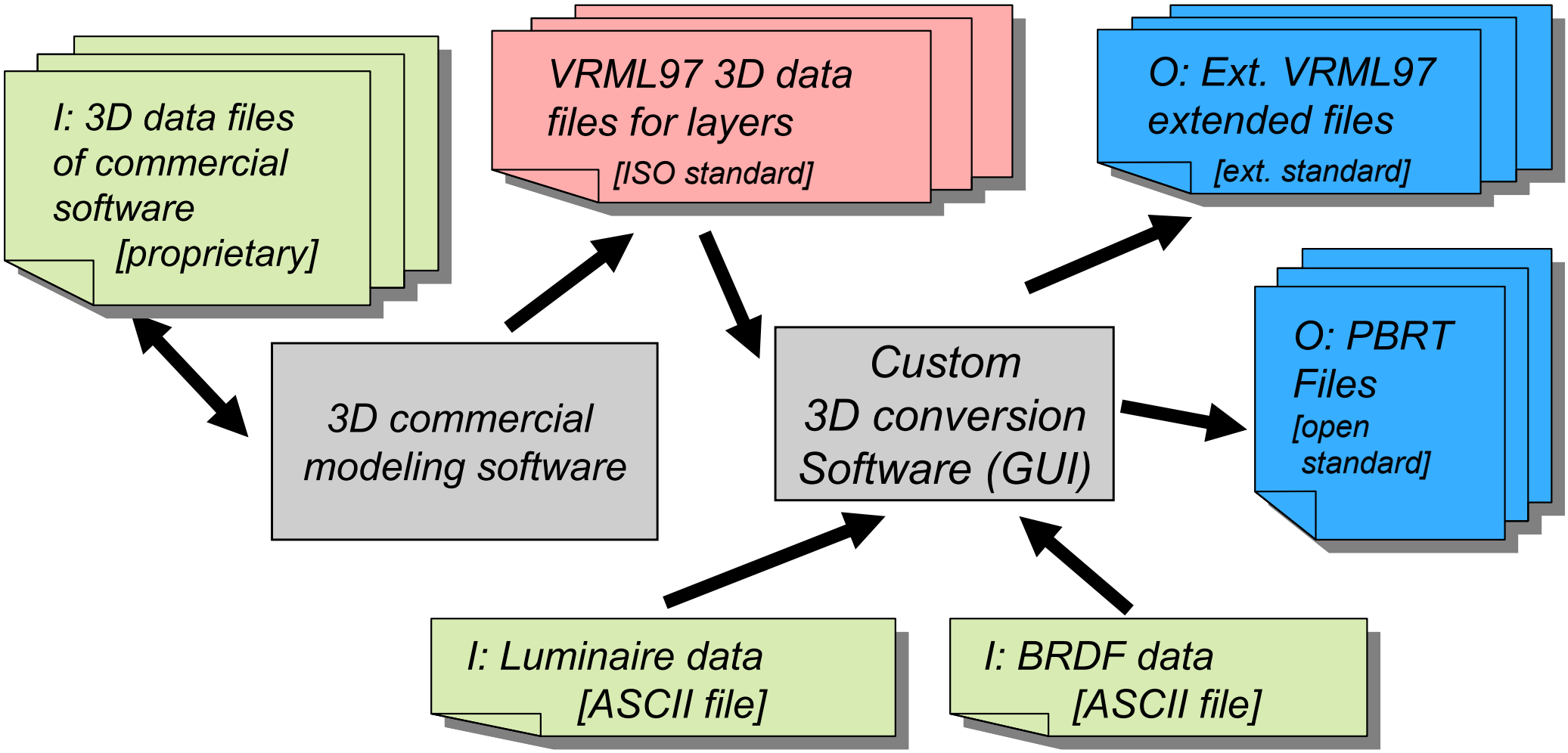




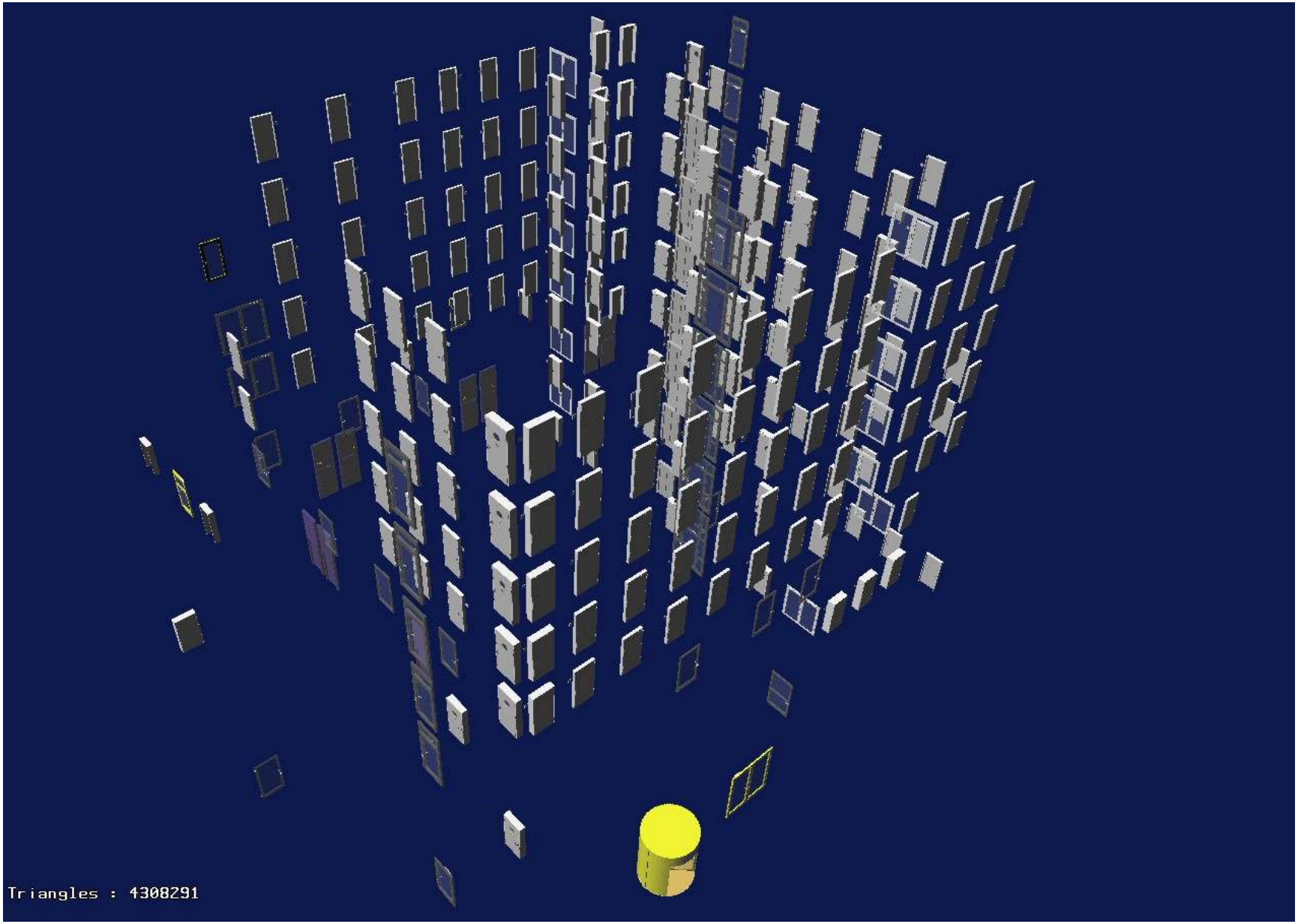


Conversion Data Flow and Processing

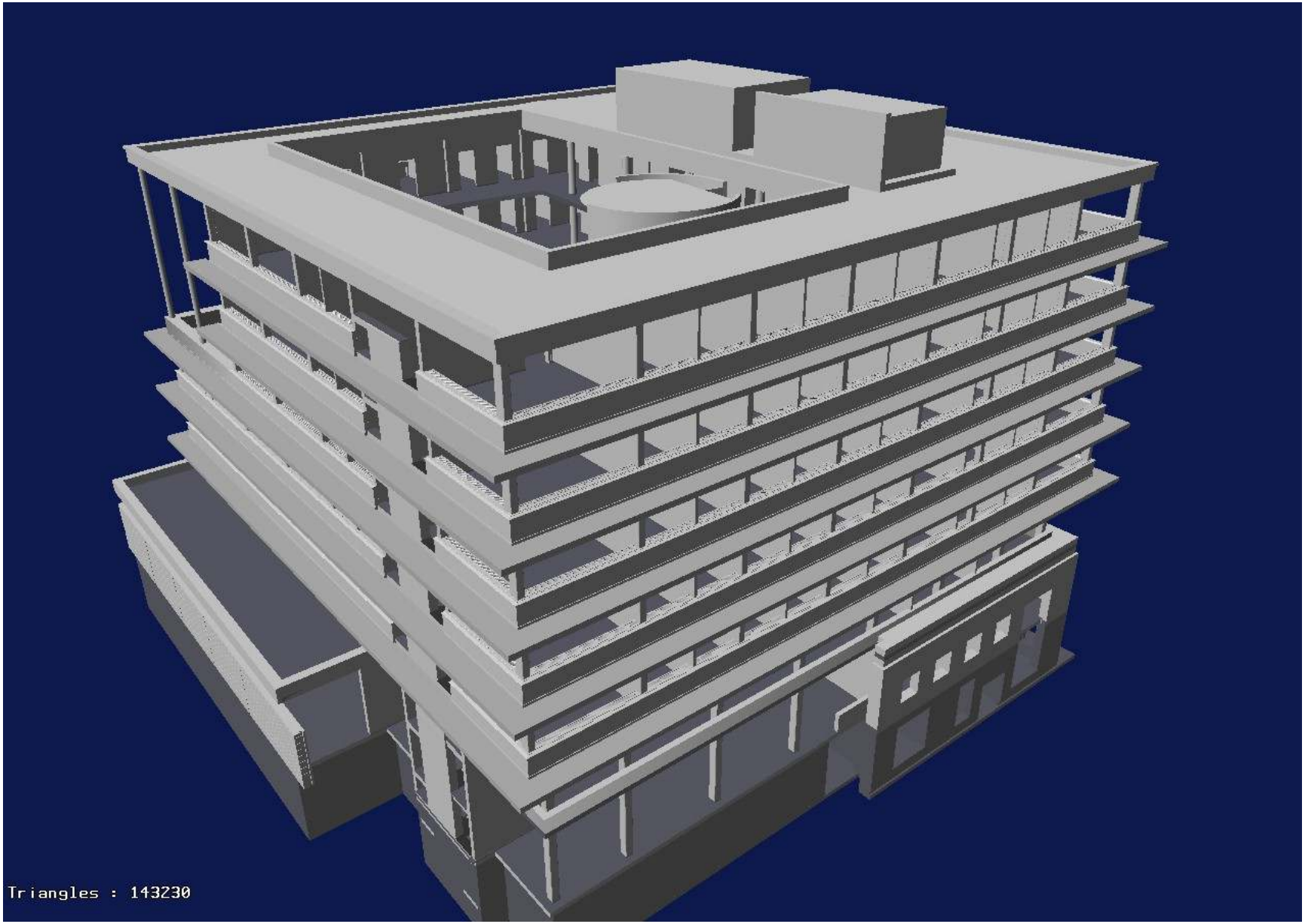
- Export feature from proprietary to open data format (geometry + luminaire + BRDF) from commercial software was missing
- Large scale 3D data – layers used. Name and data convention used in a modeling software to allow the conversion and extendibility



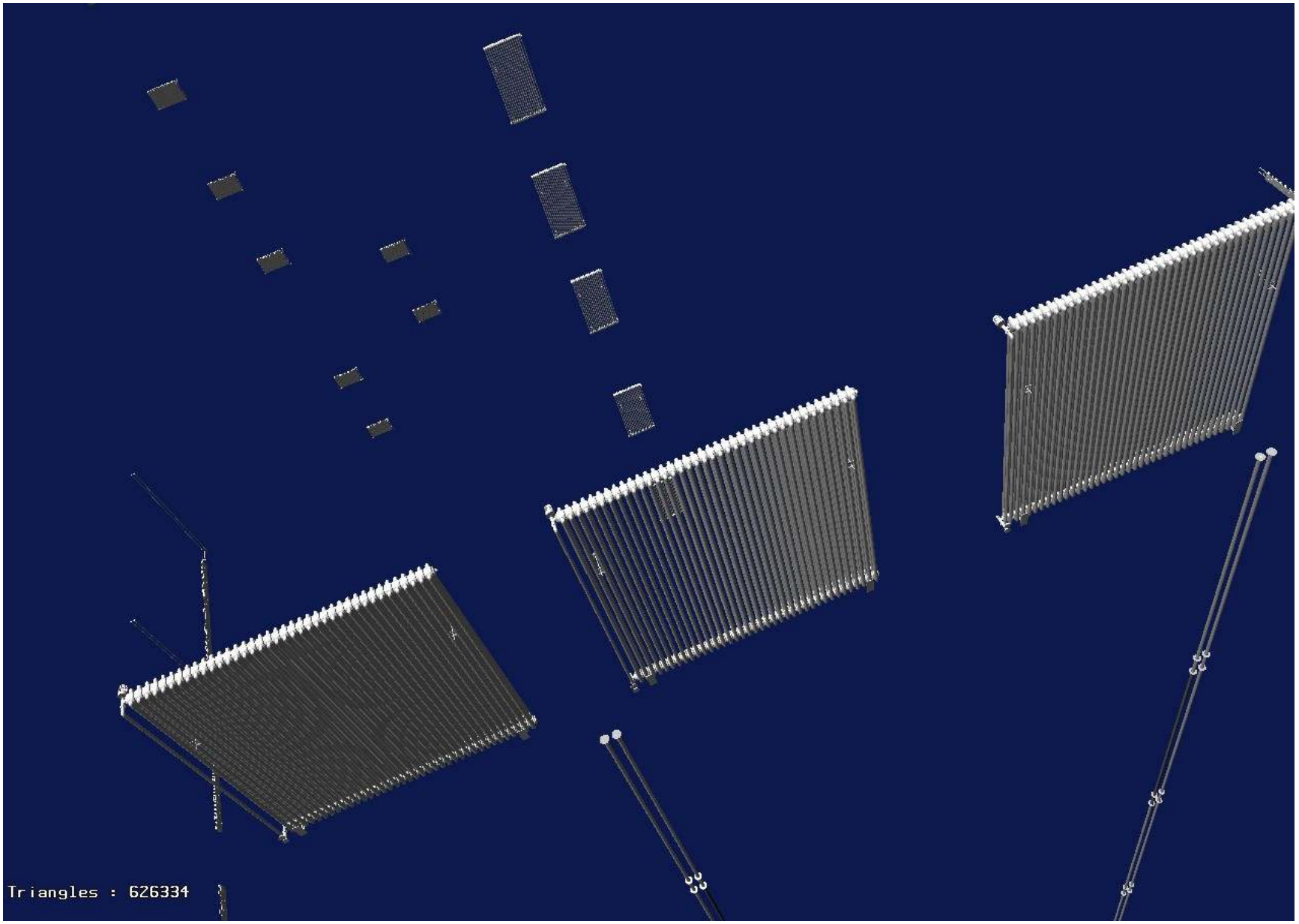
Geometry Layer Preview #1: Doors



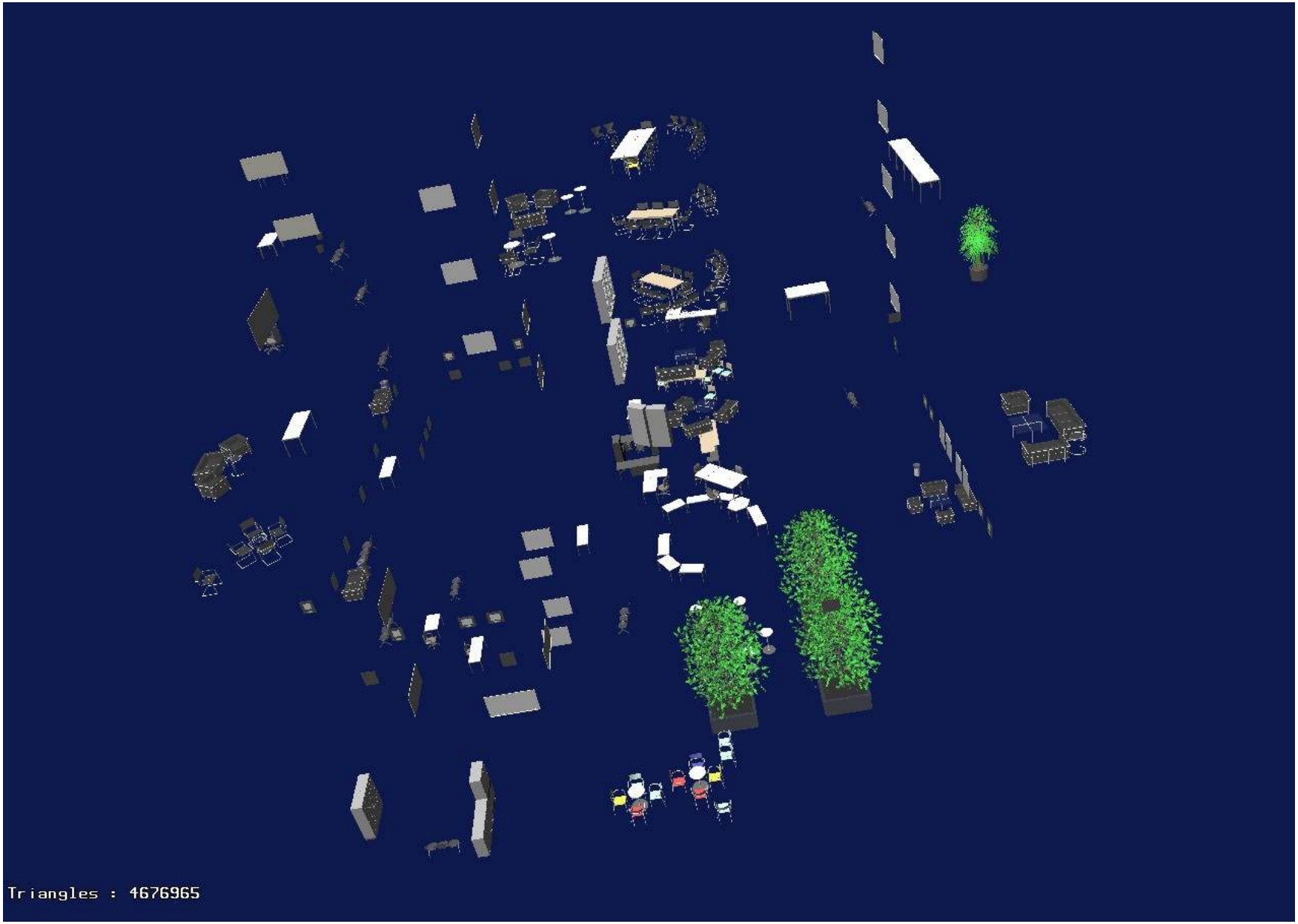
Geometry Layer Preview #2: Base Geometry



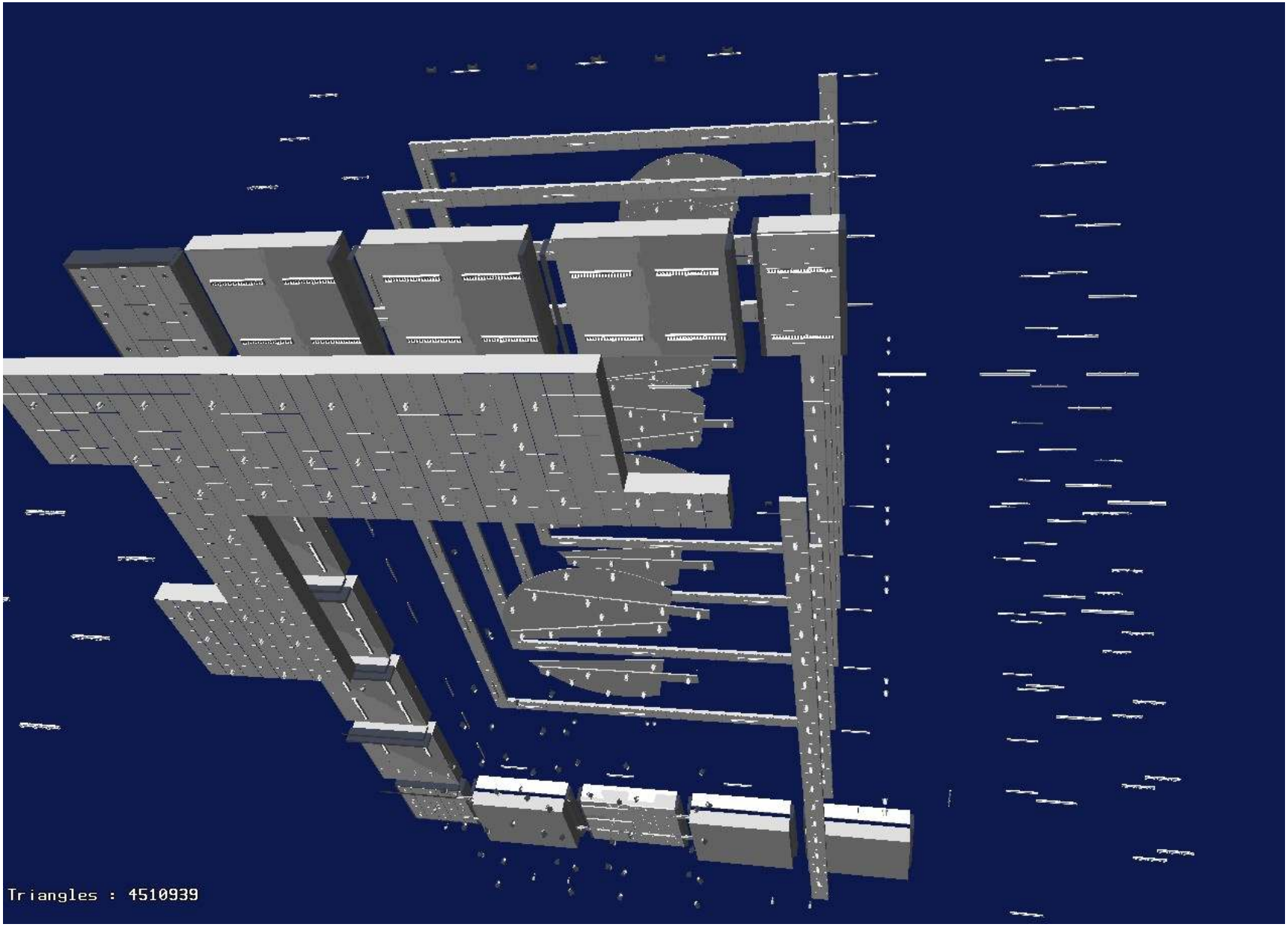
Geometry Layer Preview #3: Heating



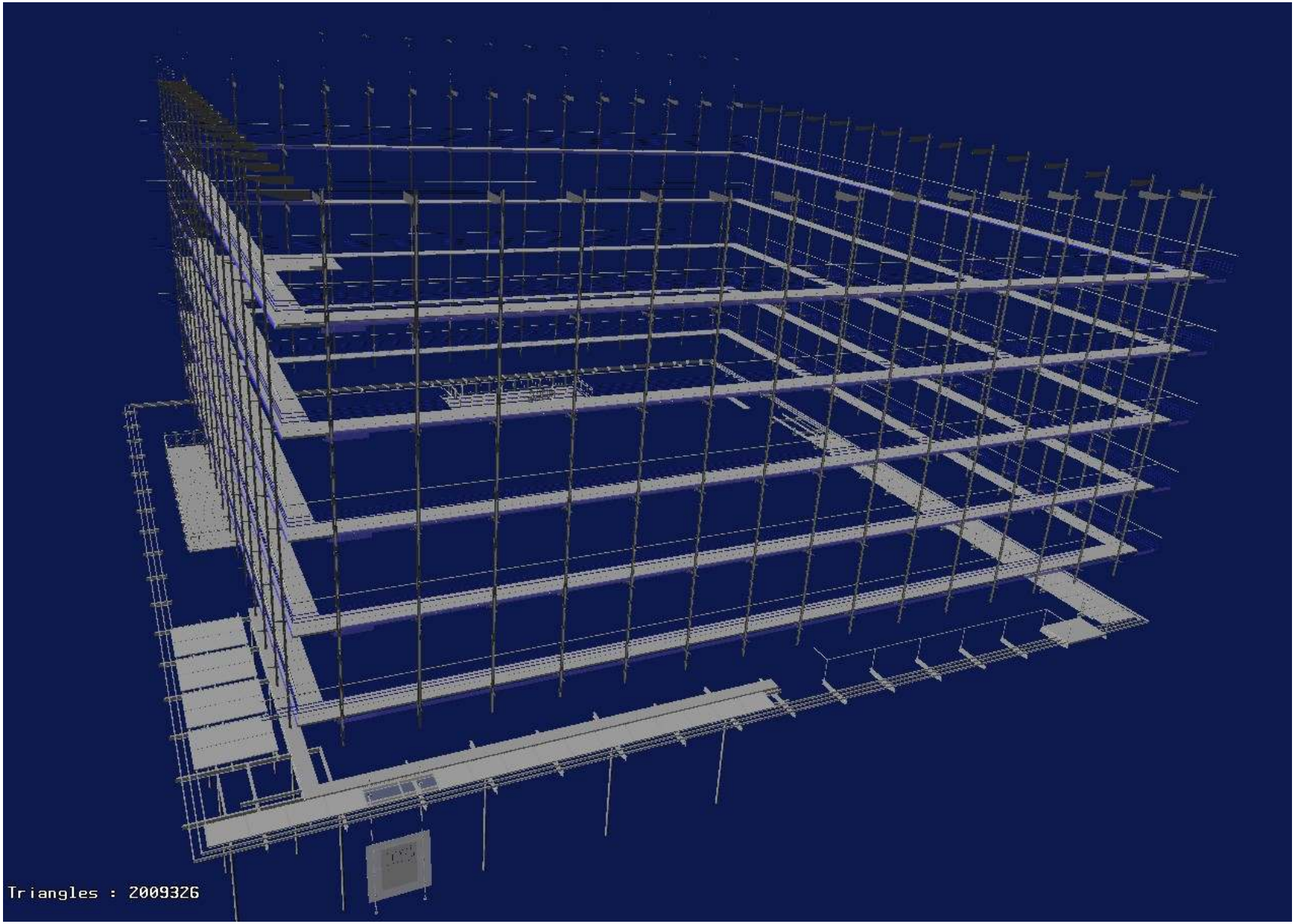
Geometry Layer Preview #4: Atrium Equipment and Furniture



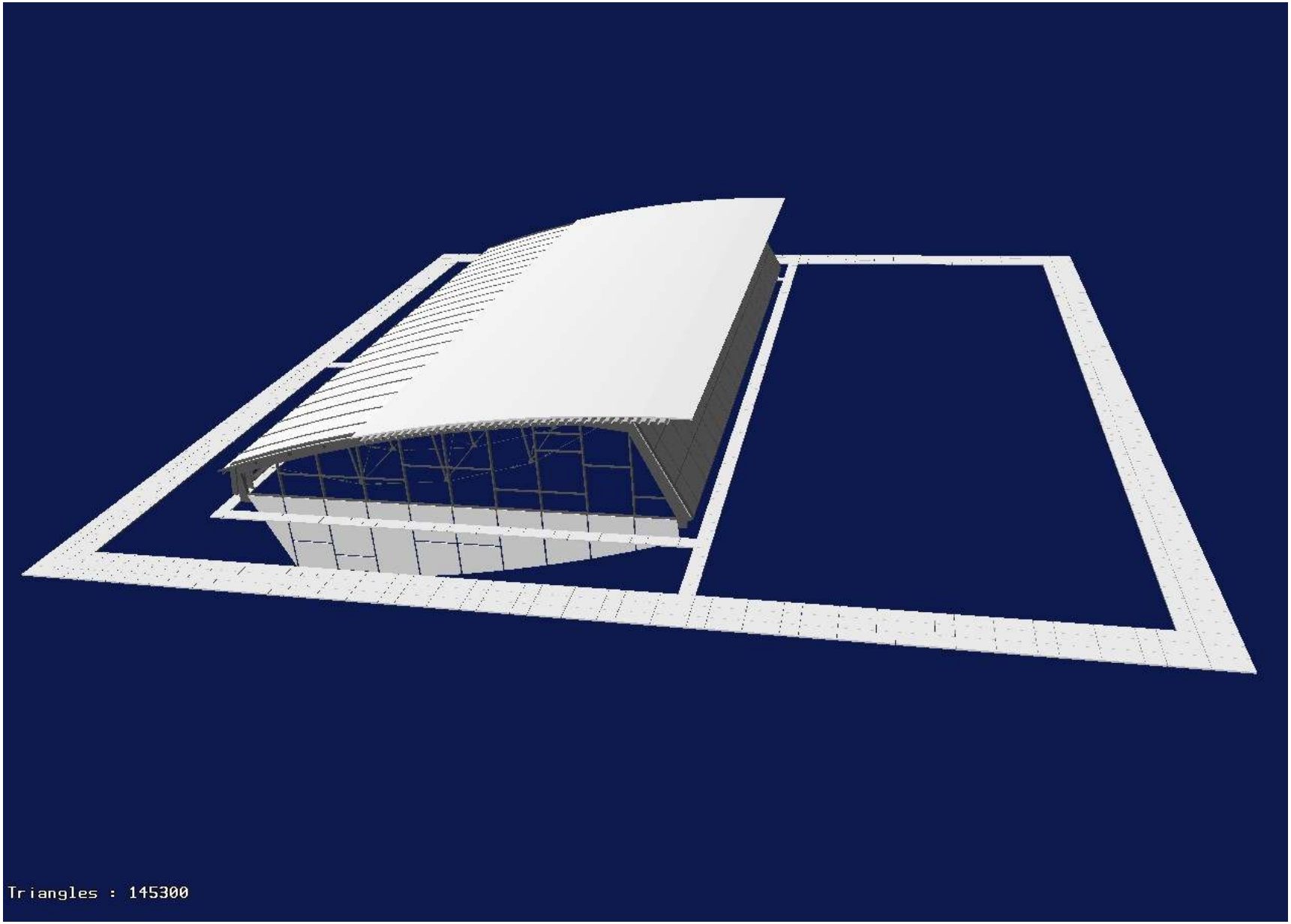
Geometry Layer Preview #5: Geometry of Luminaires



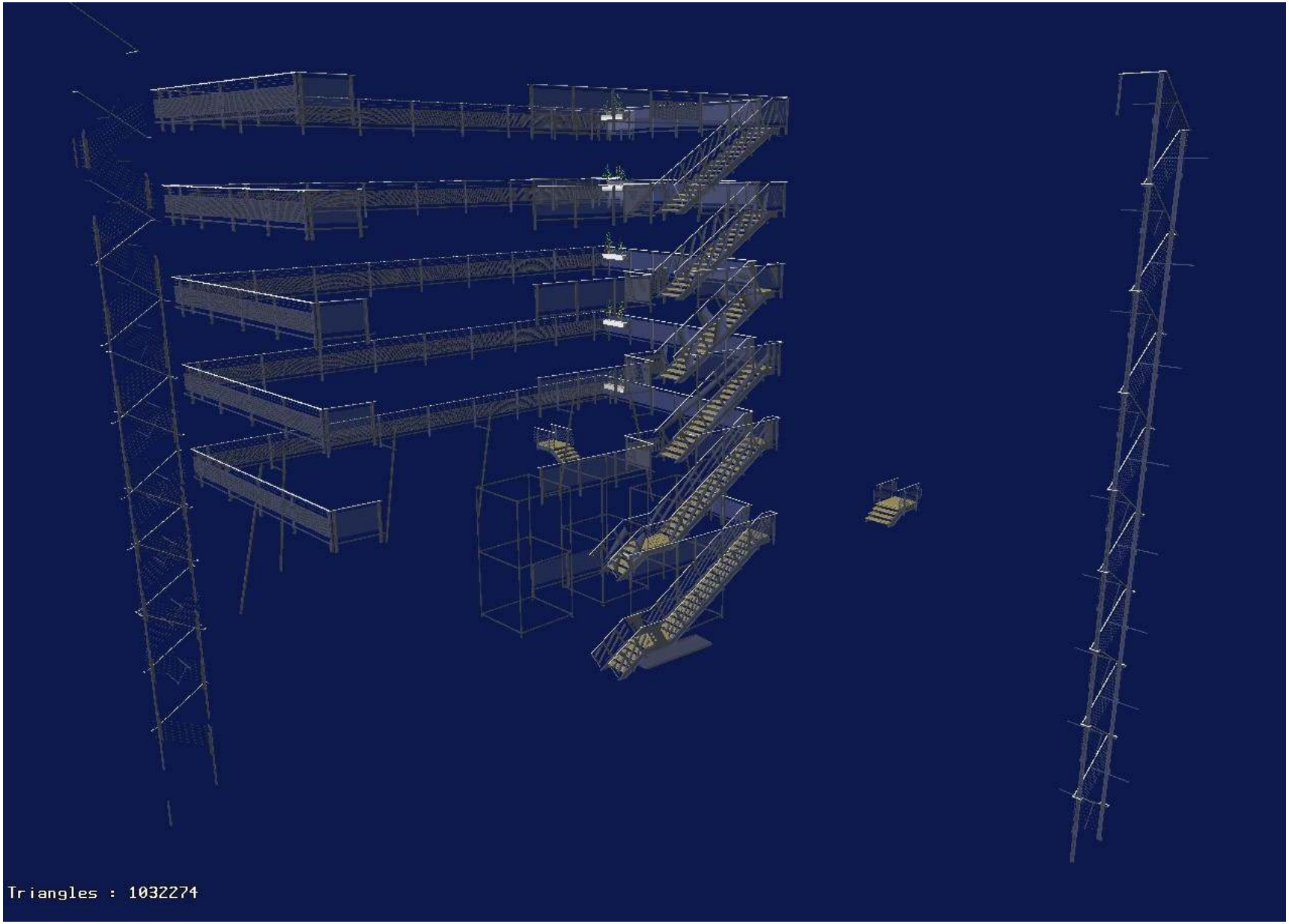
Geometry Layer Preview #6: Outer Construction



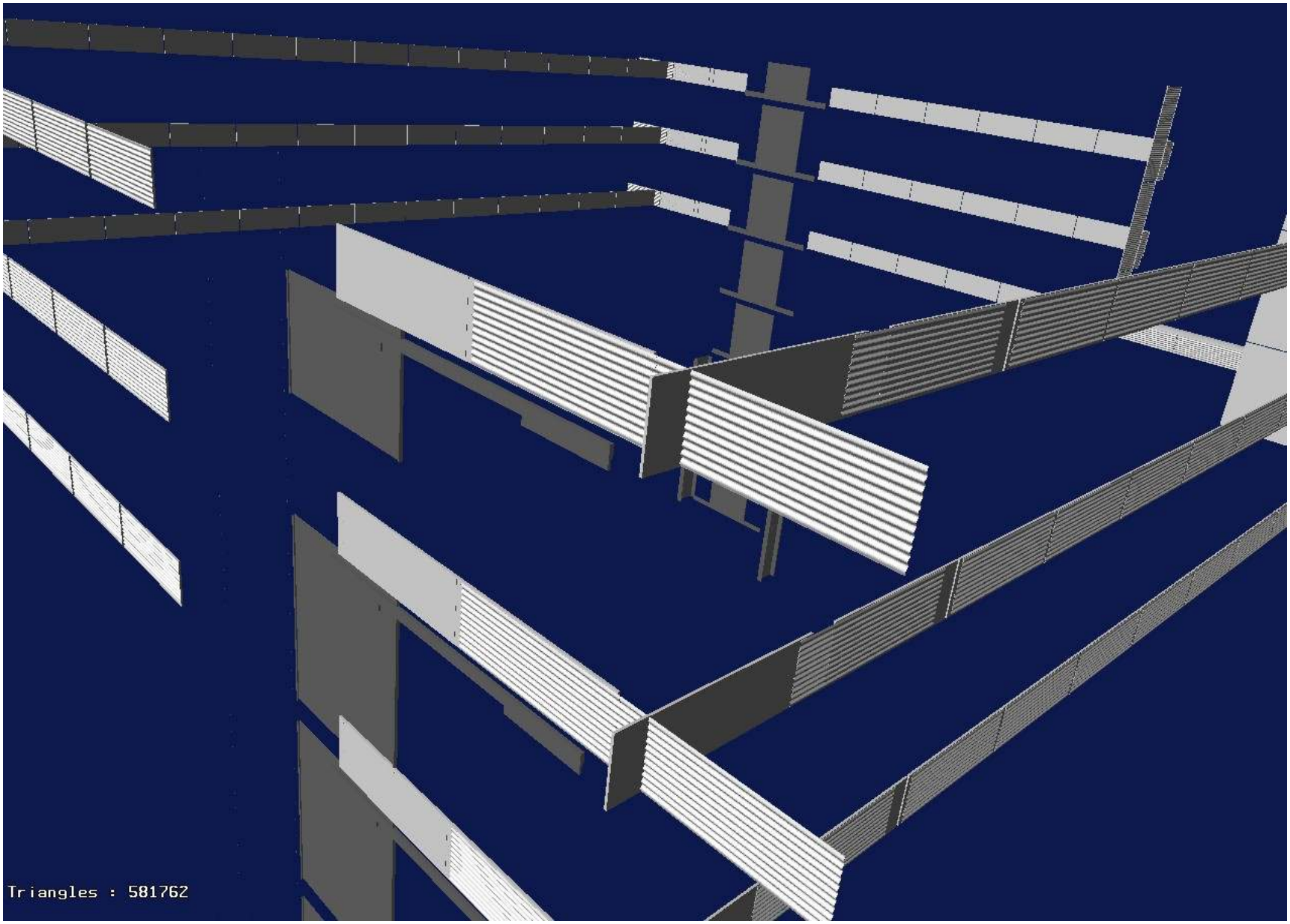
Geometry Layer Preview #7: Roof



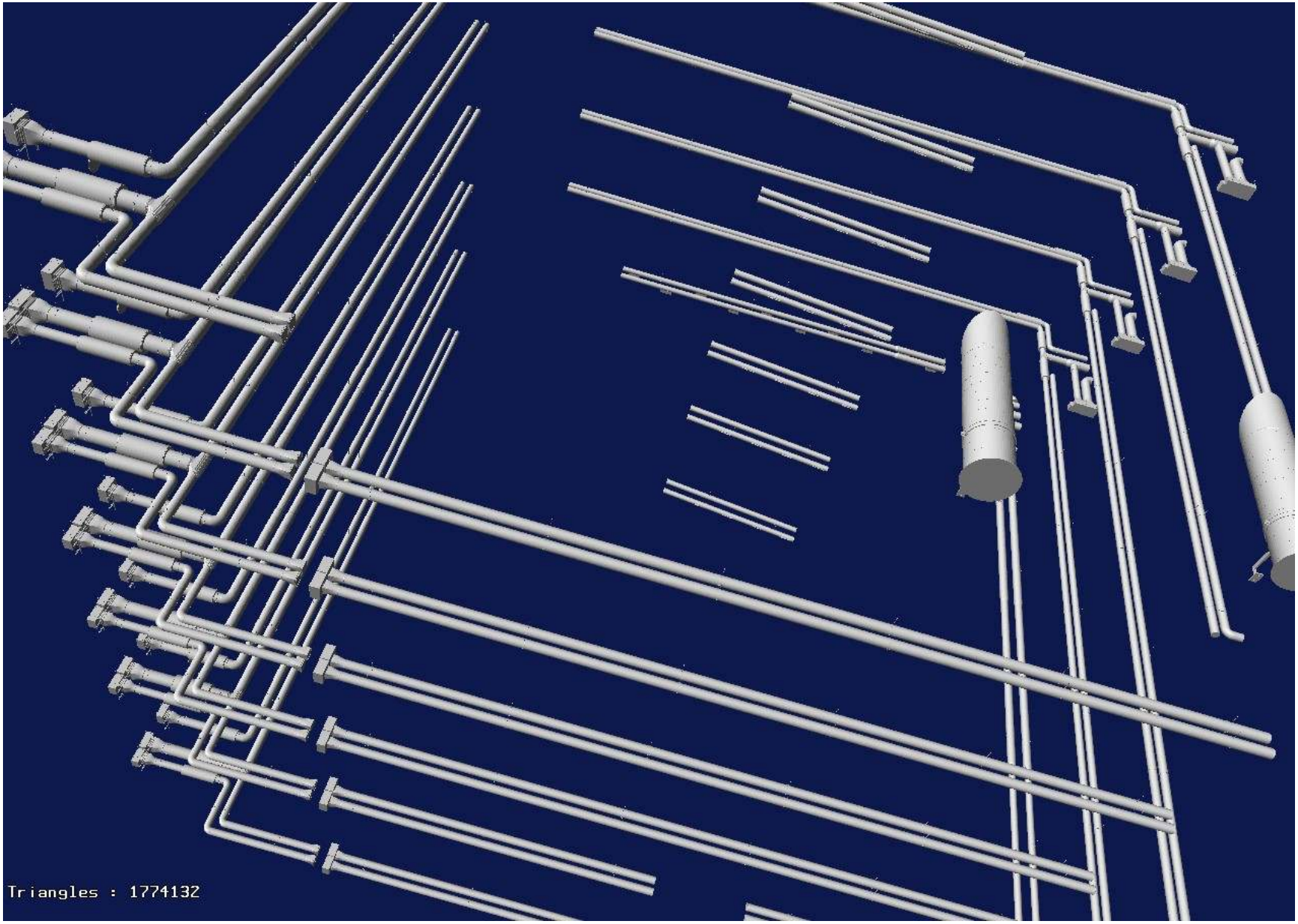
Geometry Layer Preview #8: Stairs and Fences



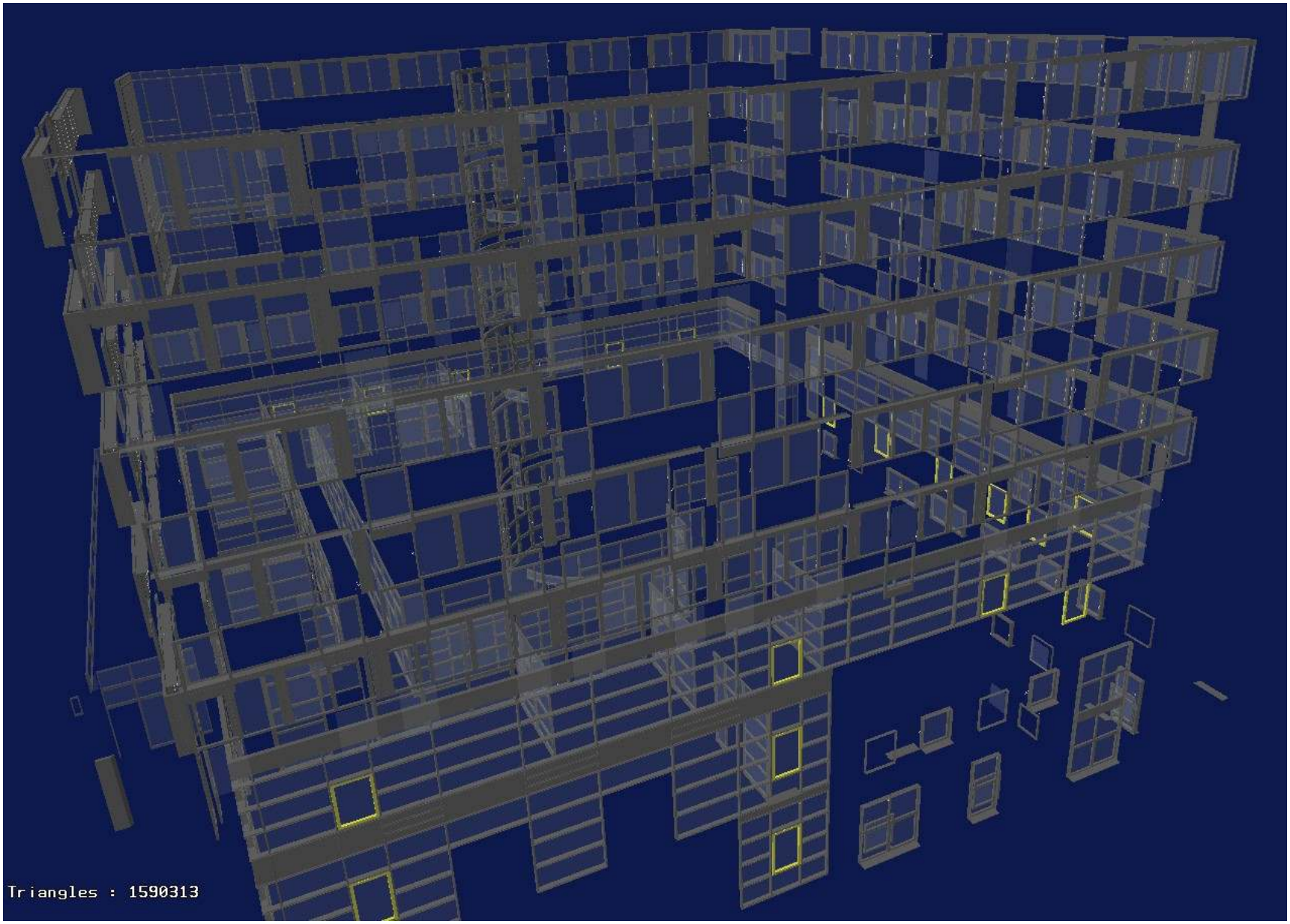
Geometry Layer Preview #9: Outer Thin Plates



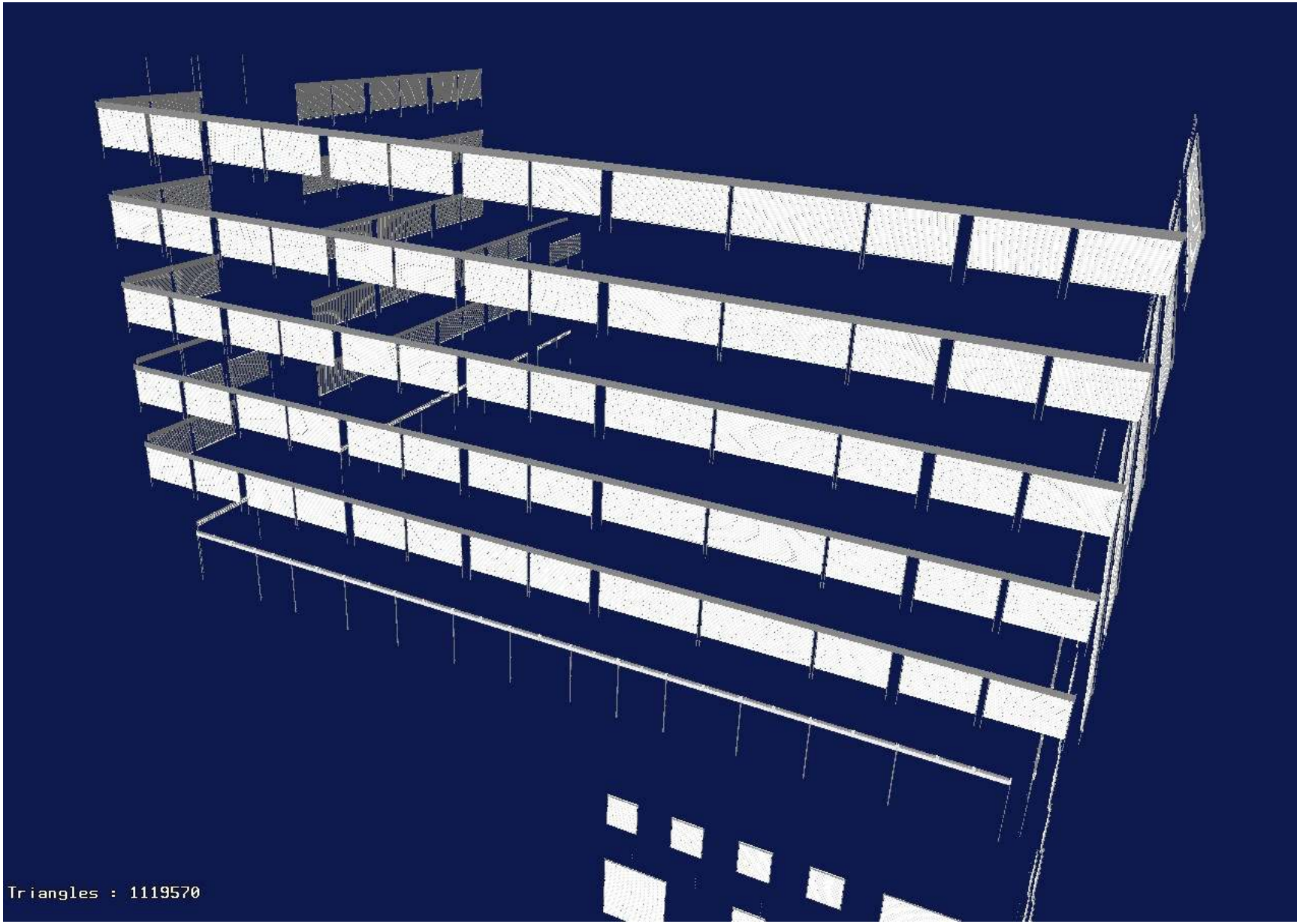
Geometry Layer Preview #10: Ventilation



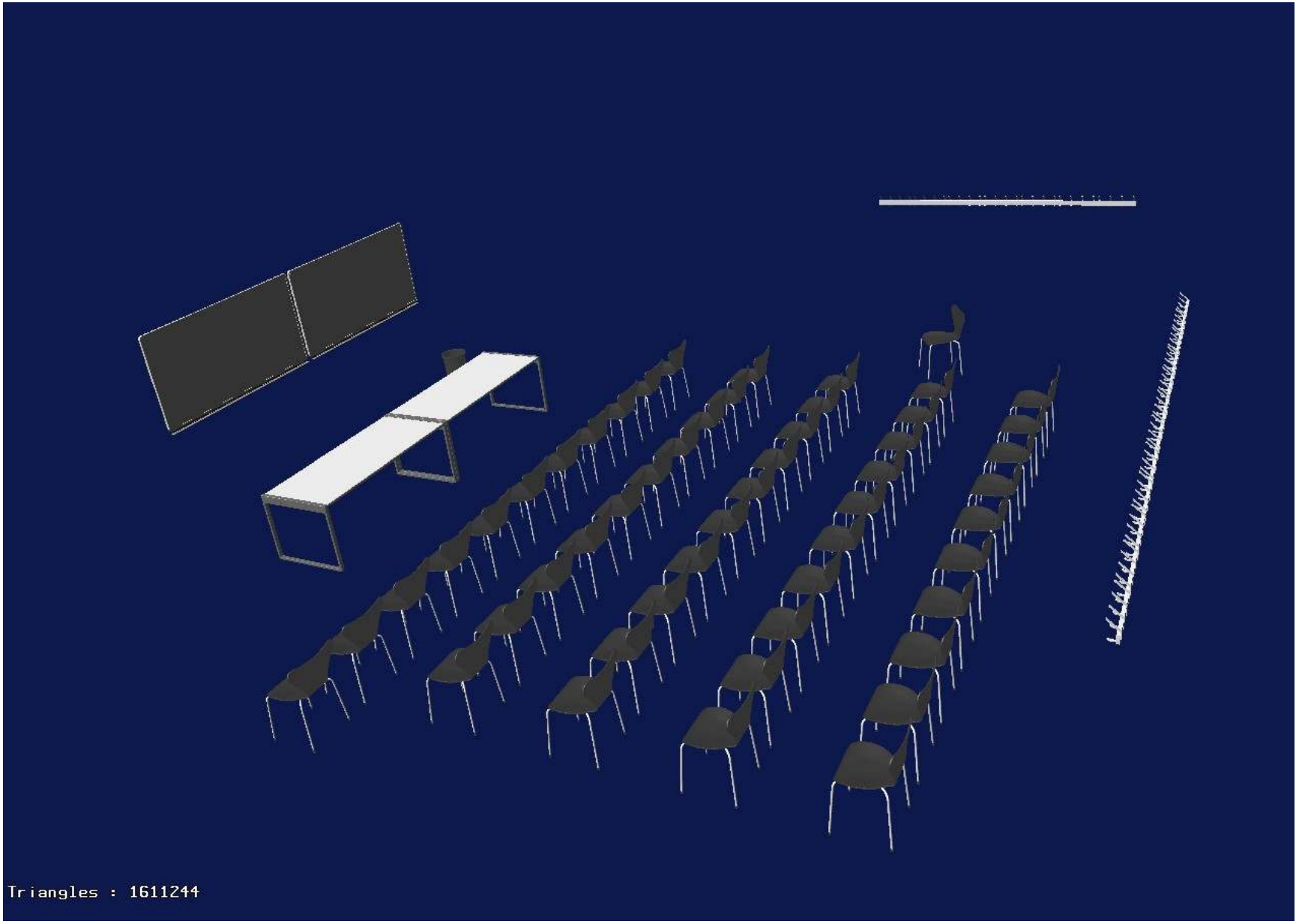
Geometry Layer Preview #11: Windows



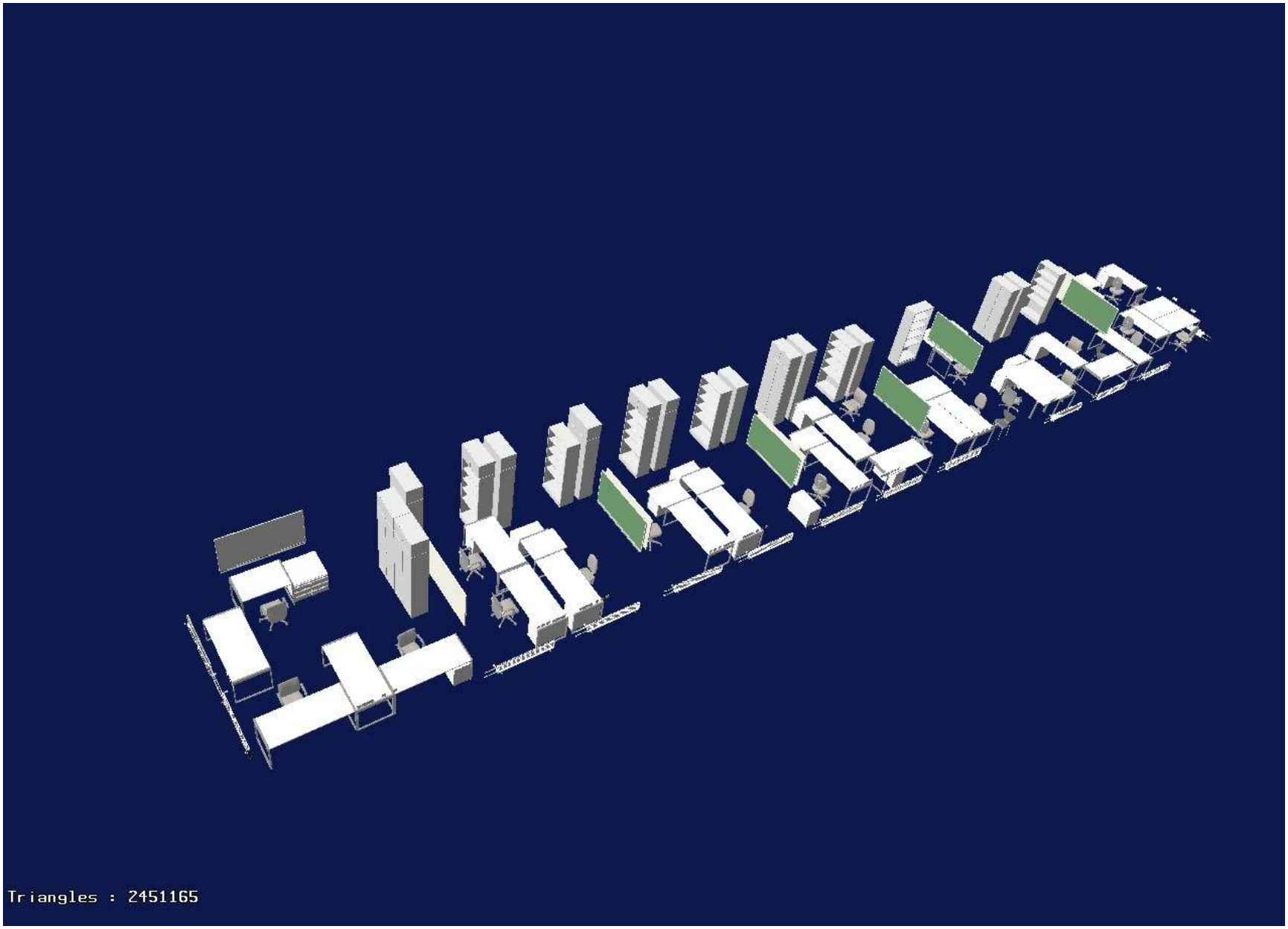
Geometry Layer Preview #12: Window Shutters



Geometry Layer Preview #13: Lecture Room 024



#14 - #34: Office Furniture, Organized to [North, East, South, West] x [2-6th Floor] + 1st floor. Example:



(Both regular and unique arrangements found.)

Comparison for Different Rendering Techniques

- PBRT software used (some problems with version 1.03)
- Selected eight views for photographs
- Three methods tested:
  - Whitted-style ray tracing, 1981
  - Path tracing by Kayija,1986
  - Instant global illumination by Wald, Kollig, Benthin, Keller, Slusallek, “Interactive Global Illumination using Fast Ray Tracing”, (Eurographics) Rendering Techniques, 2002
- Some problems found for the 3D geometry caused by the lack of backward compatibility of modeling software and wrong export of 3D data to VRML format (parts of the 3D data modeled by negative scaling)

Comparison Views (Instant Global Illumination)

View #1

View #2

Comparison Views contd.

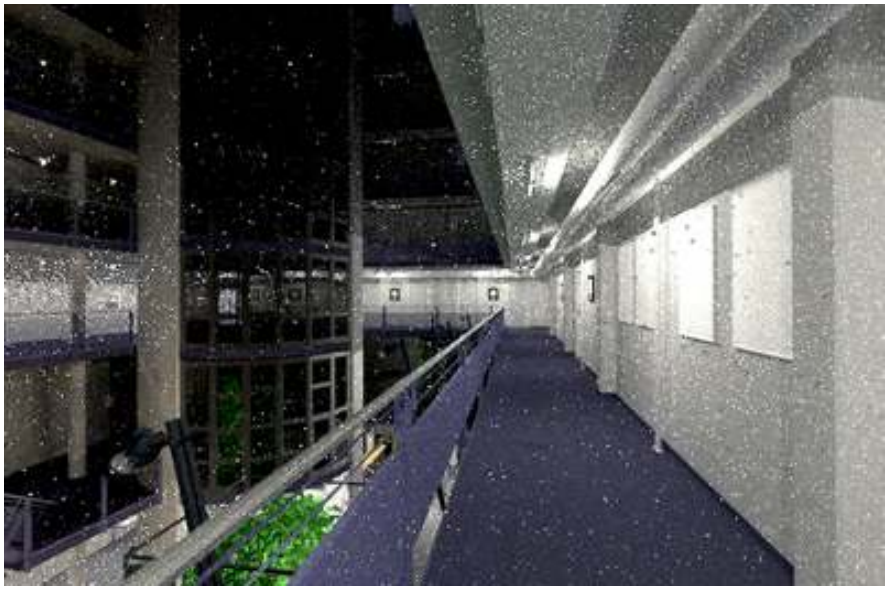
View #5

View #8

View #2 – Photo + Three Rendering Algorithms

Photograph

Path tracing



Whitted-Style Ray tracing

Instant Global Illumination

View #5 – Photo + Three Rendering Algorithms

Photograph

Path tracing

Whitted-Style Ray tracing

Instant Global Illumination

Data Model Accuracy and Limitation

Recall original project objectives (predictive rendering):

- focus on square atrium and corridor appearances
- model limited by the size of the building walls and closest features

- No parking lot and skywalk to CS building
- No equipment in the underground floor (Building Administration Department) and no showers
- No equipment and furniture on the ground floor for the administration department (rooms 002-017)
- No phone and defibrillator
- No equipment and furniture for rooms 016-023
- No furniture for south and east 1st floor (Information Services & Technology and IMPRS)

Data Model Accuracy and Limitation contd.

- No coffee machine and cups, 1st floor
- No service ladder for the roof on the 6th floor

GENERAL ITEMS

- No luminaire data in offices (emission, geometry is defined)
- No emergency lights (emission description, geometry of luminaires is defined)
- No kitchens, toilets, printer rooms, cleaning rooms
- No “Klang-und Licht Installation” by Peter Vogel
- No art exhibitions anywhere
- No furniture of group directors and secretary rooms

Data Model Accuracy contd.

- No computers and typical researcher environment of different complexities….

Data Model Accuracy contd.

- No folks…

Many thanks to Akiko Yoshida for her help with HDR photo acquisition in July 2004 !

Some Numbers for MPI Informatics Building Model

- November 2004 – **63,162** meshes (VRML97 index facets) defining **23,127,287 triangles**
- July 2009: **103,943** meshes defining **73,216,589 triangles** organized in **35** layers (allows simplification and selection)
- **623** light sources (excluding those in offices), **7** types of luminaires
- **54,595** instantiations including recursive ones
- **128,393** transformation matrices
- **53** distinct BRDFs, **0** textures, **8** HDR photos taken
- Storage needed for the webspace: approx. **19,7** GBytes
- Work spent by January 2010: approx. **4,900** hours (me, Josef, and Jiri), more than 43**0** e-mails in **4** languages, **2** visa arrangements, approx. **120** phone calls, **1** intermediate report, **1** Master thesis **(240** pages), **1** award for the Master Thesis, **1** research report (**113** pages).

Exploitation and Algorithmic Problems:

- Exploitation in computer science:
  - large scale of sizes for geometry (ratio triangle edge lengths up to 1:1000)
  - data size challenge for some algorithmic problems and majority ordinary hardware today (including recent GPUs)

- Computer Graphics
  - predictive image synthesis, lighting design, accuracy and performance test
  - real time rendering, accuracy of approximate algorithms in complex scenes
  - benchmark dataset for performance of particular hardware
  - modeling algorithms
  - reference dataset for automated reconstructions
  - environment for presentations

Exploitation and Algorithmic Problems contd.

- Computational geometry and other algorithmic problems for 3D data:
  - range searching (ray shooting, nearest neighbor search etc.)
  - robustness of geometric computations – one algorithm output:

    #number of found degenerated facets: **516,488**

    #degenerated facets to a point specified: **70,088**

    #count degenerated facets to a line segment: **445,263**

    #count degenerated facets for triangle from input file: **1,118**

    #degenerated facets for triangle as result from a triangulator = **0**

    #count zero normals from input file = **7**

    (*Note: the numbers above should be verified independently.*)
  - triangulations, remeshing
  - data conversion
  - texture parameterization

Exploitation and Algorithmic Problems contd.

- Computer Vision
  - reference data set for virtual reconstruction of a large 3D scene (rendered and for some viewpoints taken in the real building)
  - simple reference individual objects for 3D shape matching
  - technical drawings digitization
- Spatial databases algorithms
- Sound propagation algorithms
- Some others? Let me know.

Future Work Plan

- Releasing MPI Informatics Building Model V2.0
  - Correcting geometric problems
  - Parametrical model – movable geometric data (doors, windows, window shutters)
  - Acquisition of BRDF or BTF for 53 or more materials?
  - Near field photometry acquisition for luminaires?
  - Increasing complexity by further details?

Acknowledgements

- my co-workers Josef Zajac and Jiri Drahokoupil for their hard work and brevity fighting the problems and large scale complexity data.
- Hans-Peter Seidel for the project approval in 2003, partial project funding in 2003-2009, and taking the responsibility for releasing the project data on the public web in June 2010.
- Karol Myszkowski for his inspiration, advices, and remarks.
- MPII administration for many useful helps, in particular to Roxane Wetzel, Sabine Budde, Conny Liegl, Michael Bentz, and Gabi Holzer.
- MPII building administration for the access to some data.
- Michael Laise and Axel Koeppel from MPII building department for providing several spare parts from the building.
- Prof. Wolfgang Paul for his willingness to search the old 3D data on a tape, thirteen years old.
- Building architects from Kaiserslautern for providing us initial data.
- Akiko Yoshida, Prof. Jiri Habel, and Marek Balsky for their help with data acquisition.

Conclusion for MPI Informatics Building Model

- Representative publicly available real-world datasets are needed for many algorithmic problems and also required for reproducibility and the implementation validation of algorithms. Preparation of data is difficult and tedious.
- Dataset from private companies should be avoided as much as possible in research if the data cannot be made public.
- Accuracy and level of detail of 3D geometric data sets is always limited compared to the real world – we have an approximation (and MPII building is changing all the time).
- Each dataset will become obsolete in future – estimate for the MPII building model lifetime for research use is next 10 to 12 years.

Conclusion contd.

- Comments to the project are welcome, we know about some problems with the data, but do not expect / require our immediate feedback.
- Please, help to disseminate the knowledge about this project in research community.
- Project data are intended for the research purposes, please, use it if they fit to your projects. We do not want that our work is wasted. Thank you.
- Any support for this project is really appreciated.
- The project website is located at:

http://www.mpi-inf.mpg.de/resources/mpimodel/v1.0