

UNDERSTANDING HUMAN PERCEPTION OF BUILDING CATEGORIES IN VIRTUAL 3D CITIES - A USER STUDY

P. Tutzauer^{a,*}, S. Becker^a, T. Niese^b, O. Deussen^b, D. Fritsch^a

^a Institute for Photogrammetry, University of Stuttgart, Germany – forename.lastname@ifp.uni-stuttgart.de

^b Dept. of Computer and Information Science, University of Konstanz, Germany - forename.lastname@uni-konstanz.de

ThS16: Perceptual and cognitive experiments with imagery and 3D models

KEY WORDS: Urban Modeling, Scene Understanding, Building Classification, 3D, Perception, User Study

ABSTRACT:

Virtual 3D cities are becoming increasingly important as a means of visually communicating diverse urban-related information. To get a deeper understanding of a human's cognitive experience of virtual 3D cities, this paper presents a user study on the human ability to perceive building categories (e.g. residential home, office building, building with shops etc.) from geometric 3D building representations. The study reveals various dependencies between geometric properties of the 3D representations and the perceptibility of the building categories. Knowledge about which geometries are relevant, helpful or obstructive for perceiving a specific building category is derived. The importance and usability of such knowledge is demonstrated based on a perception-guided 3D building abstraction process.

1. INTRODUCTION

This work presents first research results of project D01 'Perception-Guided Adaptive Modeling of 3D Virtual Cities' which contributes to SFB/Transregio 161 'Quantitative Methods for Visual Computing'. Virtual 3D cities are the basis for a growing number of applications. Besides providing *geometric* information on the represented buildings, virtual 3D cities can also serve as medium to visually communicate urban- or building-related *semantic* information.

The degree of insight that people obtain via this visual communication strongly depends on what kind of geometric 3D building representations are used. Geometric 3D representations which fit people's visual habits and urban legibility can help to achieve a quick and accurate understanding of urban spatial information. Virtual 3D cities can be based on various data types and ways of modeling (e.g. unstructured 3D point clouds, meshed surfaces, textured or non-textured volumetric 3D models with different levels of detail and abstraction) (Fritsch et al., 2011; Haala, 2013; Mayer et al., 2012; Kolbe et al., 2005; Gröger and Plümer, 2012). However, the question 'Which of these geometric 3D representations is, given a context, best suited to enable a maximum understanding of the information that is intended to be transmitted?' is still an open problem.

This paper serves as basis to target the question of which kind of geometric 3D representation will enable the user to gain the required degree of insight for a specific task by gathering profound knowledge on the human's ability to understand semantics from 3D building structures.

Generally, it depends on the application as to which specific degree building-related semantic information needs to be understood by the user. Semantic issues of interest may be: building category, architectural style, historical relevance, state of preservation etc. Out of these, we will exemplarily address the semantic issue 'building category' which covers basic semantic information: Being able to quickly understand the category of buildings when moving through virtual 3D cities means support for various applications (e.g. navigation, house hunting, real estate management, spatial marketing) as it will

help users to orient themselves and enable intuitive and efficient exploration.

Within the paper, we present a user study we developed and conducted in order to reveal the required knowledge about how a human understands building categories from geometric 3D building representations. We will identify geometric properties and structures which are relevant for perceiving specific building categories. Moreover, we will demonstrate how such knowledge about perceptually relevant geometric structures can be applied to improve the interpretability of 3D building abstractions.

The paper is structured as follows: Section 2 describes the development and conduction of the user study. Results of the test as well as a first application of the derived knowledge are shown in section 3. The paper ends with conclusions and an outlook in section 4.

2. DEVELOPMENT AND CONDUCTION OF THE USER STUDY

Our study was designed to investigate different aspects of how different types of building representations affect the user's decision of classifying a building into a certain category. Analyses are expected to provide knowledge to create the best application-dependent building representation that can be interpreted intuitively and thus enabling the user to achieve a comprehensive understanding of building-related semantic information. Test persons had to classify buildings into six building categories: **One-Family Building** (OFB), **Multi-Family Building** (MFB), **Residential Tower** (RT), **Building With Shops** (optionally with partial residential usage) (BWS), **Office Building** (OFF), **Industrial Facility** (IF). The buildings which are to be classified are randomly taken from German cities (mostly Stuttgart), i.e., between 15 and 20 candidates of each building category were selected and LoD3 models manually modeled. For 60% of the buildings, additionally, textured meshes/LoD2 models from Google Earth and/or images from Google Street View were provided. Figure 1 gives

* Corresponding author

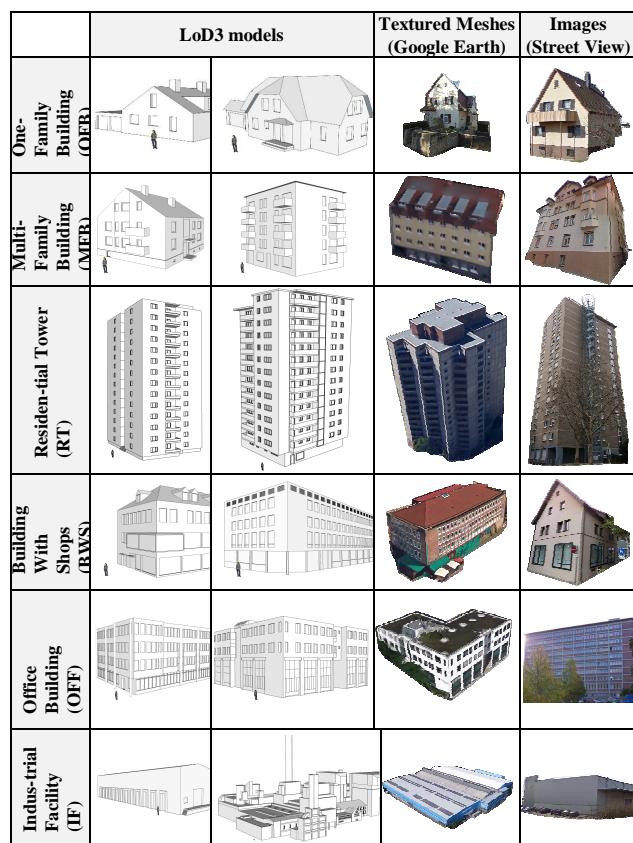


Figure 1. Examples for building categories and representation types used in the study (Google Earth/Street View, ©2015 Google)

examples of the building categories and representation types presented to the user.

Some general information about the test persons were obtained beforehand: Gender, Age, Graduation, Subject of study, Nationality and whether or not previous experiences in 3D virtual reality worlds (computer games, Google Earth, CAD modeling etc.) is existent. In total, 165 different building representations had to be classified in random order. Additionally, for each representation, users had to rate their level of certainty (reaching from ‘Very Uncertain’ to ‘Very Certain’ in 5 selection options). Based on this self-assessment, a relation between user correctness and certainty can be examined. This metric can give further information about whether the user is aware of being wrong in the current classification.

2.1 Evaluation Metrics

The actual reference category for each model is obtained by extracting the type of use from the digital city basic chart and 3D data from the City Surveying Office of Stuttgart. To compare differences between the user’s classification and the actual ground truth, all surveys are evaluated, and typical classification quantities such as confusion matrix, commission/omission errors and user’s/producer’s accuracy are computed. Moreover, in order to obtain deeper knowledge on the user’s perception, for each building category, the ground truth buildings are compared to the classified buildings. Aiming at quantifiable results, this comparison is based on computing geometric building properties inherent in LoD3 models. Following properties are evaluated: building footprint, number of floors, floor height, total building height, number of windows per façade, mean window surface area, window-to-wall-surface

ratio, number of entrances, mean entrance surface area, number of balconies, mean balcony surface area, relative frequency of different roof types, different appearance of ground floor (GF) compared to remaining floors. Regarding the latter, 4 different aspects are analysed: different arrangement, size and shape of windows in GF, as well as different ground plan in GF than in other floors. Each of these 4 aspects can take either the value 1 (different) or 0 (equal). Thus, the 4 mean values, which are computed for all representatives of a building category, express the degree of geometric difference between ground floor and remaining floors.

3. RESULTS AND FIRST APPLICATION

Overall, 96 test persons have participated in the user study. In the following, we will first evaluate the classification results based on the entirety of all users (section 3.1). Concrete knowledge of the users’ perception of building categories is derived in section 3.2. A first application of the obtained knowledge, namely perception-based abstraction, is presented in section 3.3.

3.1 Evaluation Based on the Entirety of All Users

Afore-mentioned classification quantities are computed to reveal differences between the user’s classification and the actual ground truth. The results are shown in Table 1.

	Producer Accuracy [%]	User Accuracy [%]	Commission Error [%]	Omission Error [%]
OFB	73.6	91.8	8.2	26.4
MFB	73.3	68.3	31.7	26.7
RT	73.3	77.8	22.2	26.7
BWS	62.7	64.8	35.2	37.3
OFF	71.2	64.4	35.6	28.8
IF	94.3	90.2	9.8	5.7

Table 1. Classification metrics obtained from confusion matrix

Obviously, *OFB* and *IF* could be identified best with both over 90 percent user accuracy. Users have most difficulties with the classes *OFF*, *BWS* and *MFB* which is indicated by user accuracies between 64.4% and 68.3%. Reasons for that will be further explained in section 3.2. Besides the classification result, users should also rate their certainty for each particular decision. For 22 buildings the correct classification result was below 50%, with a mean correctness of 32.4% for these buildings. However, the mean certainty value for the same buildings is 3.78, which translates to a certainty level of closely to ‘Certain’. This reflects the issue of the user often not even knowing of currently misinterpreting data. Even more: users might feel certain in their wrong classification. Therefore, it is necessary to use derived knowledge about the difference between perception/expectation and reality to optimize the building representation for the user’s needs.

3.1.1 Evaluation Based on Different Groups of Users

In the following, we analyzed whether different groups of users come to different classification results. The participants of the study have been quite homogeneous with respect to *age* (90% between 18 and 30 years, with a total range from 18 to 73), *graduation* (over 90% higher education entrance qualification, Bachelor or Master), and *subject of study* (over 95% engineering studies). However, clearly separable user groups of meaningful size can be identified with respect to *gender* (71% male, 29% female), the users’ *origin* (38.5% German, 61.5%

		Ø Footprint [m ²]	# Floors	Floor Height [m]	Total Height [m]	# Windows Per Facade	Ø Window Surface Area [m ²]	Window/Wall Surface Ratio [%]	# Entrances	Ø Entrance Surface Area [m ²]	# Balconies	Ø Mean Balcony Surface Area [m ²]	Different Window Arrangement in GF	Different Window Sizes in GF	Different Window Shapes in GF	Different Ground Plan in GF	Roof Complexity
Ground Truth	OFB	115.30	2.1	3.30	9.58	8.3	1.33	16.1	1.7	4.13	0.2	1.04	0.67	0.27	0.33	0.13	2.7
	MFB	238.41	4.0	2.94	14.68	27.8	1.93	30.0	1.4	2.99	2.2	2.06	0.18	0.18	0.18	0.06	2.7
	RT	573.01	15.5	2.73	43.67	110.8	1.94	29.1	1.2	3.53	9.5	4.24	0.36	0.36	0.36	0.21	1.0
	BWS	697.51	3.7	3.84	17.61	34.9	3.58	52.2	1.6	24.90	1.1	10.08	0.91	0.91	0.82	0.55	2.3
	OFF	868.26	5.9	4.09	24.94	96.8	4.94	127.9	1.4	7.93	0.1	3.42	0.53	0.41	0.41	0.53	1.0
	IF	10812.6	2.5	24.63	59.60	23.0	10.17	10.1	5.3	13.33	0.0	0.00	0.53	0.47	0.47	0.40	1.4
As Classified	OFB	148.69	3.1	8.65	18.25	22.2	3.47	25.9	1.6	4.99	0.9	1.80	0.53	0.30	0.30	0.20	2.3
	MFB	252.92	5.8	5.43	22.98	48.7	2.31	32.5	2.2	7.35	2.6	3.82	0.49	0.38	0.40	0.29	1.9
	RT	328.95	6.9	3.55	24.19	61.0	3.84	56.9	1.4	8.39	3.0	4.36	0.46	0.39	0.38	0.28	1.8
	BWS	432.80	5.7	6.78	26.69	56.5	4.52	53.6	2.3	9.86	2.0	3.30	0.54	0.47	0.46	0.36	1.8
	OFF	480.00	6.3	6.61	28.11	59.9	3.84	52.5	2.3	8.97	2.1	3.34	0.49	0.44	0.41	0.32	1.7
	IF	912.6	5.0	11.10	36.78	59.3	6.19	61.7	2.8	13.48	1.5	3.88	0.53	0.47	0.49	0.40	1.4

Table 2. Geometric properties of the building categories as given in the ground truth (upper part of the table), and as classified by the users in the study (lower part of the table)

foreign - from all over the world) as well as *previous experience* with 3D virtual reality worlds (75% experience, 25% none).

Thus, the user study is additionally evaluated with respect to the latter three properties. For this purpose, the same accuracy measures as in section 3.1 have been determined, this time, however, for the different user groups separately. Student's t-tests were carried out to search for significant differences in the classification results between those user groups. However, no significant differences in the classification results of the aforementioned user groups could be identified, therefore all subsequent evaluations and interpretations in section 3.2 are based on the entirety of all participants.

3.2 Derivation of Knowledge on Building Perception

Based on the findings described in section 3.1, we derived coherences between the perceptibility of the building categories and several properties of the 3D representations. In section 3.2.1, *geometric dependencies*, i.e., dependencies between the perceptibility of a building's category and the building's geometric properties are extracted. In section 3.2.2, the perceptibility with respect to different representation types is analyzed.

3.2.1 Perceptually Relevant Building Structures

To derive geometric building properties and structures which are relevant or essential for the perceptibility of a specific building category, we first analyzed the geometric properties of the building categories' representatives of our ground truth (see section 3.2.1.1). Afterwards, the same analysis was done for building categories as perceived by the users (see section 3.2.1.2).

3.2.1.1 Metrics of Building Categories Reference: As results of the investigations on geometric properties of ground truth data, following facts can be noted:

- *One-Family Buildings*: significantly smaller in terms of footprint, number of floors and total height in comparison to all other categories.
- *Multi-Family Buildings*: resemble *Buildings With Shops* most. Important difference: *MFB* have a smaller mean window surface area. Additionally, window arrangement, size and shape as well as ground plan on ground floor are very similar to the remaining floors.

- *Residential Towers*: high amount of floors, balconies and windows as well as total height are essential.
- *Building With Shops*: appearance of the ground floor significantly differs from the remaining floors. This holds true for both the window size and the arrangement of windows.
- To correctly recognize *Office Buildings* and distinguish from *BWS*, a higher number of windows per façade as well as a higher amount of floors is characteristic. Accordingly, the ground floor and first floor resemble each other more in contrary to *BWS*.
- *Industrial Facilities*: feature the biggest footprint overall. Whereas *IF* have typically a small number of floors, in return, however, a great floor height.

3.2.1.2 Metrics of Building Categories as Perceived/Expected by Users: The same mean features are computed based on the total amount of buildings all users classified into the respective class (see Table 2(lower part)). To compare ground truth data with results from all users, a significance test for the differences in all corresponding features is computed – this way discrepancies in the user's perception or expectation and ground truth can be revealed.

The most important findings in this evaluation are:

- *One-Family Buildings*: no difference in perception and ground truth.
- A different arrangement of windows on the ground floor as well as a different ground floor itself in comparison to the remaining floors of the buildings is expected for *Multi-Family Buildings*. Additionally, in the users' perception *MFB* have a higher number of floors.
- For users, the number of floors can be less and the total height lower in comparison to ground truth, to classify a building as *Residential Tower*. However, a single floor height is expected to be higher than for the ground truth.
- *Buildings With Shops*: considered to have a higher number of floors than in reality.
- *Office Buildings*: users are expecting a higher number of balconies.
- *Industrial Facilities*: expected to have more windows per façade and a bigger number of floors.

3.2.2 Findings Based on Building Representation Type

By separating the evaluation into geometric and textural representation types, their impact onto the classification results can be measured. The mean correctness for untextured LoD3 models is at 69.2%. Whereas a slightly higher correctness could be achieved for the textured meshes/LoD2 models from Google Earth with 75.4%. However, the most accurate classification result with 79.3% is based on the images from Google Street View. The difference between untextured LoD3 models and textured meshes/LoD2 models from Google Earth is not significant but there is a significant difference between the geometric representation and images from Street View. One reason for the superior correctness obtained for Street View representations could be the viewpoint of the models. As exemplarily shown in the last column of Figure 1, all images are captured looking slightly upwards and thus resembling the human perspective.

For building categories that are easily separable from the rest like *One-Family Buildings* and *Industrial Facilities*, a geometric representation is sufficient in the majority of cases. Particularly for buildings that are belonging to a somewhat more ambiguous categories like *BWS*, *OFF* and *MFB* additional textural information improves the classification results.

3.3 First Application: Perception-Based Abstraction

The knowledge derived in section 3.2.1 describes geometric 3D building properties and structures which are characteristic for a specific building category. This knowledge can now be used for perception-based building abstractions. Due to the vast amount of mobile devices, virtual 3D cities have to be abstracted and geometrically simplified for display in applications on small screens. Thereby it is important that the abstracted building representations still contain those geometric properties and structures which are essential for perceiving the correct category of a building. The following example shows how such perceptual knowledge can be embedded in a 3D abstraction process and effects on the perceptibility of the buildings' categories are demonstrated. Based on Nan et al. (2011), we created different abstractions of buildings based on human perception. During this process, we use the Gestalt laws of proximity, regularity and similarity to group blocks together and represented the results by larger blocks. The preservation of geometric properties and 3D structures, which are essential for perceiving the correct building category, is ensured by translating them into geometric constraints as restrictions for the abstraction process. Figure 2(a) depicts a model belonging to the class of *Building With Shops*. The first abstraction (a2) incorporates the properties learned to be important for *BWS* as mentioned in section 3.2.1.1. The ratio of the window size between the ground floor, the first floor and the remaining floors is preserved as well as the arrangement of the windows. The second model is a free abstraction. As a result of the abstraction process, both models (a2) and (a3) have merged dormers. However, the window shapes and distribution have changed. For example (a2) retains smaller windows in the upper floor, while (a3) has a merged window front. In Figure 2(b) a *Residential Tower* is depicted. For both abstractions, windows have been merged over two floors, as a consequence the total building height appears to be smaller and the number of floors decreases with increasing single floor height at the same time. This exactly corresponds to the findings made for the users' expectation of the category *Residential Tower* (see Table 2). The important feature 'balcony' is maintained in the first abstraction, the second abstraction however drops it. This way

model (b2) retains the essential appearance, whereas the appearance of model (b3) diverges more from an actual RT.

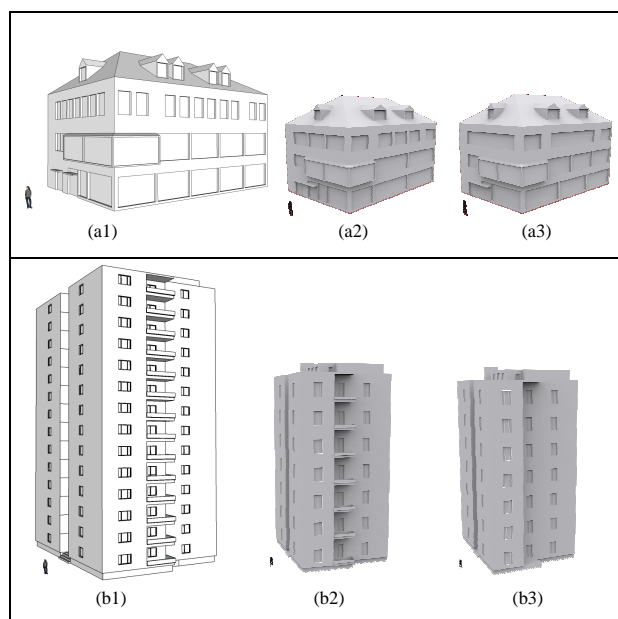


Figure 2. First application for conclusions drawn from the survey. For $x \in \{a, b\}$: (x1) original building model, (x2) abstraction based on features important for the user to classify into the respective correct category, (x3) 'free' abstraction without restrictions

4. CONCLUSIONS

With the aim of deriving some first knowledge on the human's ability to understand semantics from 3D building structures, we presented a user study on the user's comprehension of building categories based on different 3D building representations. Analyses of the user study reveal clear coherences and dependencies between the correctness of classifications and the model representation type. In general it is conducive to have textural information for buildings. Users tend to experience difficulties to distinguish between *Buildings With Shops*, *Multi-Family* and *Office Buildings*. Beyond, the majority of the users is not even aware of their misinterpretations which makes perception-adapted building representations an even more important issue. Therefore, it is crucial to guide the representation based on features that are significantly characteristic for the respective building category. As a first application, we demonstrated how such knowledge about the human's perception of building-related semantic information can be used for the perceptually adapted abstraction of 3D building models.

ACKNOWLEDGEMENTS

We would like to thank the German Research Foundation (DFG) for financial support within the projects D01 and A04 of SFB/Transregio 161. Additionally, we would like to thank the European Social Fund (ESF) as well as the Ministry Of Science, Research and the Arts Baden-Württemberg for financial support within the 'Margarete von Wrangell-Habilitationsprogramm für Frauen'.

REFERENCES

- Fritsch, D., Khosravani, A., Cefalu, A., Wenzel, K., 2011. *Multi-sensors and multiray reconstruction for digital preservation*, in: Fritsch, D. (Ed.), Proceedings of the Photogrammetric Week. Wichmann, pp. 305–323.
- Gröger, G., Plümer, L., 2012. *CityGML - Interoperable semantic 3D city models*. Int. J. Photogramm. Remote Sens. 71, 12–33.
- Haala, N., 2013. *The landscape of dense image matching algorithms*, in: Fritsch, D. (Ed.), Proceedings of the Photogrammetric Week. Wichmann, pp. 271–284.
- Kolbe, T.H., Gröger, G., Plümer, L., 2005. *CityGML - interoperable access to 3D city models*, in: Oosterom, Zlatanova, Fendel (Eds.), Proceedings of the International Symposium on Geoinformation for Disaster Management. Springer, pp. 883–899.
- Mayer, H., Bartelsen, J., Hirschmüller, H., Kuhn, A., 2012. *Dense 3D reconstruction from wide baseline image sets*, in: Proceedings of the International Conference on Theoretical Foundations of Computer Vision. Springer-Verlag, Berlin, Heidelberg, pp. 285–304.
- Nan, L., Sharf, A., Ke, X., Deussen, O., Cohen-Or, D., Chen, B., 2011. *Conjoining Gestalt rules for abstraction of architectural drawings*. ACM Trans. Graph. 30, 185:1–185:10.