PFG 2016 / 5–6, 319–333
Stuttgart, December 2016

Article

A Study of the Human Comprehension of Building Categories Based on Different 3D Building Representations

Patrick Tutzauer, Susanne Becker, Dieter Fritsch, Stuttgart, Till Niese & Oliver Deussen, Konstanz

Keywords: user study, building categories, urban modeling, scene understanding, human perception

Summary: Virtual 3D cities are becoming increasingly important as a means of visually communicating diverse urban-related information. Since humans are the direct recipients of this information transfer, it is vital that the 3D city representations account for the humans' spatial cognition. Thus, our long-term goal is providing a model for the effective perception-aware visual communication of urban- or building-related semantic information via geometric 3D building representations which induce a maximum degree of perceptual insight in the user's mind. A first step towards this goal is to get a deeper understanding of a human's cognitive experience of virtual 3D cities. In this context, the 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.

Zusammenfassung: *Eine Studie über die menschliche Wahrnehmung von Gebäudekategorien auf Basis unterschiedlicher 3D-Gebäuderepräsentationen.* Virtuelle 3D-Städte werden zunehmend wichtig, um unterschiedlichste stadtrelevante Informationen visuell zu vermitteln. Da Menschen die direkten Empfänger dieses Informationstransfers sind, ist es unerlässlich, dass 3D-Stadtrepräsentationen die räumliche Wahrnehmung von uns Menschen berücksichtigen. Unser längerfristiges Ziel ist es daher, ein Modell zur wahrnehmungsbewussten visuellen Kommunikation von städte- oder gebäudespezifischen semantischen Informationen zu entwickeln, welches über geometrische 3D-Gebäuderepräsentationen dem Nutzer ein Maximum an Erkenntnisgewinn ermöglicht. Ein erster Schritt dorthin ist, sich ein besseres Verständnis der menschlichen Wahrnehmung von virtuellen 3D-Städten zu verschaffen. In diesem Zusammenhang präsentiert der Beitrag einen Nutzertest über die menschliche Fähigkeit, Gebäudekategorien (z.B. Wohngebäude, Büros, Gebäude mit Läden usw.) anhand geometrischer 3D-Gebäuderepräsentationen zu erkennen. Die Studie zeigt zahlreiche Abhängigkeiten zwischen geometrischen Eigenschaften der 3D-Repräsentationen und der Wahrnehmbarkeit der Gebäudekategorien auf. Wissen darüber, welche geometrischen Eigenschaften relevant, hilfreich oder hinderlich sind, um eine bestimmte Gebäudekategorie zu erkennen, wird aus den Ergebnissen der Studie abgeleitet. Die Wichtigkeit und der Nutzen dieser Erkenntnisse werden anhand einer wahrnehmungsgesteuerten Abstraktion von 3D-Gebäudemodellen aufgezeigt.

DOI: 10.1127/pfg/2016/0302 1432-8364/16/0302 $ 3.75

1 Introduction

Virtual 3D cities are used in a growing number of applications: They are the basis for decision makers in areas such as urban planning, policy making for environmental aspects or planning for evacuation and emergency response. Moreover, 3D city models have also entered people's everyday life in the meantime via 3D navigation and tourist information systems or computer games and augmented reality 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. In this case, the 3D representations should enable the users to fast and intuitively comprehend the respective semantics without wasting mental workload on non-relevant information. The degree of insight that people obtain via the visual communication of semantics 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. Due to the multitude of different sensors, algorithms and modeling concepts used for acquiring and processing geodata in urban areas, 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. 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.

Depending on the application and the requirements going along with it, the provision of virtual 3D cities may involve considerable investments with respect to costs, time and expertise for data acquisition and processing. Thus, it is highly unsatisfactory that it is not known beforehand whether the desired degree of understanding can be reached by means of the generated virtual 3D building representations, or whether a smaller solution would have been sufficient. Questions like these are of special relevance for developers of systems that work with 3D virtual cities, e.g. 3D navigation systems, virtual reality applications, computer games etc. The overall goal of this project is to provide a tool that can be used by developers of such systems to determine which kind of geometric 3D representation will enable the user to gain the required degree of insight: The tool will allow to quantify, predict and enhance the degree of perceptual insight induced by specific 3D building representations in a specific context. However, the basis for all that – profound knowledge on the human's ability to understand semantics from 3D building structures – is still missing.

This paper provides an important first step towards the project's overall goal by dealing with the identification of perceptual aspects which are relevant for the understanding of semantic information inherent in geometric 3D building structures.

Generally, it depends on the application as to which specific 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 will present a user study which we developed and conducted in order to reveal the required knowledge about how a human understands building categories from geometric 3D building representations. In more detail, we will focus on two questions:

1. Which representation type is for which building category the most suitable?
2. Which geometric building properties and structures are relevant for the perceptibility of a particular building category?

Moreover, we will demonstrate how the derived 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 gives an overview of related work. The development and conduction of the user study is described in section 3. Section 4 shows results of the test as well as an application of the derived knowledge. The paper ends with conclusions and an outlook in section 5.

2 State of the Art

Without raising claim to completeness, we briefly comment on geometric representation types used for virtual 3D cities in section 2.1. Related work on the human perception of geometric building structures is given in section 2.2 while section 2.3 addresses research on the quantification of perceptual aspects.

2.1 *Geometric Representations of Virtual 3D Cities*

The variety of geometric representations of urban scenes is wide: Most virtual 3D cities are a collection of 3D buildings given as boundary representations (BReps). Following CityGML, the OGC standard for 3D city models (Kolbe et al. 2005, Gröger & Plümer 2012), the geometric level of detail (LoD) of 3D building representations can range from LoD1 and LoD2 (LoD1: box models using flat roofs, LoD2: detailed roof structures, planar façades), which are available for the majority of the buildings of a 3D virtual city – over LoD3 (3D façade structures), which are usually only available for single landmarks and small test scenes – up to LoD4 (indoor models), which are not within the scope of our project.

Due to increasing computing power, nowadays, urban scenes can also be represented based on dense unstructured 3D point clouds or triangle meshes. These models are either the direct output of laser scanning or, pushed by the development of Structure-from-Motion and dense multi-image matching techniques (Hirschmüller 2008, Agarwall et al. 2009, Engel et al. 2014), the result of photogrammetric derivation from images (Fritsch et al. 2011, Haala 2013, Mayer et al. 2012). Google Earth, for example, solely uses triangle meshes for their representations. By this, they avoid the derivation of geometrically and possibly also semantically interpreted BReps with a defined LoD which, however, are required for all applications that go beyond pure visualizations.

2.2 *Human Perception of Geometric Building Structures*

Research on the human perception of 2D geometric objects stems from a variety of different branches of science, e.g. geoinformatics and photogrammetry, geography, cartography or computer graphics. Findings of Gestalt theory play an important role in this. For example, Li et al. (2004) exploit Gestalt principles for the grouping and generalization of 2D building footprints, and Michaelsen et al. (2012) refer to Gestalt-based groupings for the detection of 2D window structures in terrestrial thermal imagery. Within the wide field of visualization approaches, Adabala et al. (2009) present a perception-based technique for generating abstract 2D renderings of building façades, and Nan et al. (2011) apply conjoining Gestalt rules for the abstraction of architectural 2D drawings.

Approaches on the human perception of geometric building representations, which are not restricted to 2D structures or 2D visualizations but, instead, are directly located in 3D space, are often developed in the context of cartography. In this context, most approaches aim at the reduction of the visual complexity of urban 3D representations to decrease the user's cognitive effort. Prominent representatives are provided by Glander & Döllner (2009) or Pasewaldt et al. (2014), who use cognitive principles for generating abstract interactive visualizations of virtual 3D city models. Both approaches focus on emphasizing landmarks while buildings that are supposed to be unimportant from a tourist's point of view are grouped and replaced by cell blocks. Instead of using Gestalt rules, this grouping is based on the infrastructure network. Other approaches realize the abstraction of virtual 3D cities by directly analyzing and modifying the geometric properties of the building models. For example, Sun et al. (2011) propose a structure-preserving abstraction method which generates abstracted 3D building models by avoiding concave shapes.

All the works mentioned have one thing in common: They integrate perceptual principles in their methods for the recognition, generalization or abstraction of geometric building structures in order to reveal or emphasize building-related information. These perception-based methods, however, are all more qualitative than quantitative operations. That means, quantitative statements about the degree to which the respective information can be perceived by a human, or tasks like, for example, searching for the best abstraction to achieve a certain degree of perceptibility are not supported.

2.3 *Quantifying Human Perception of Geometries*

Existing attempts to quantify the human perception of geometric objects are closely linked to Gestalt principles and, therefore, limited to simple 2D structures. Desolneux et al. (2004) and Cao et al. (2007) propose a probability measure to quantify the meaningfulness of groupings in cluster analysis for 2D shape recognition. Kubovy & van den Berg (2008) provide a probabilistic model of Gestalt based groupings by proximity and similarity on regular 2D patterns. Michaelsen & Yashina (2013) put the Gestalt principles in an algebraic setting to facilitate 2D object recognition in images.

To the best of our knowledge, the evaluation of complex 3D building geometries with respect to their perceivable semantic information content, i.e. the quantification of perceptual insight, has not been addressed yet. Based on our user study on the human perception of building categories (see the following sections), we will take a first step in this direction.

3 Development and Conduction of the User Study

The overall goal of this user study is to obtain knowledge about the user's comprehension of building categories in virtual 3D cities. In more detail, the study is 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 answers to questions such as 'Which representation type is for which building category the best?' or 'Which geometric building properties and structures are relevant for the perceptibility of a particular building category?'. Knowledge like that can be of great benefit when – given a specific application – the task is to provide the best suitable building representations which can be interpreted most intuitively, and, thus, enable the user to achieve a quick and correct understanding of building-related semantic information. Within a 3D navigation tool for example it is not crucial to provide the highest level of detail, since users should be able to identify essential structures with a glimpse. While Virtual Reality applications, as its name implies, aim at preferably detailed representations.

The data basis and the setup of the user study are described in section 3.1; the applied evaluation metrics are presented in section 3.2. The results of the study as well as a first application scenario showing how the derived knowledge can be used for perception-aware abstraction processes will be part of section 4.

3.1 *Data Basis and Setup*

The category of a building is reflected in both *geometric* building properties, e.g. building size, roof shape, size, number and arrangement of windows etc., and textural information. In order to separate the influences of both aspects as well as possible, the following representation types are used within the study: (a) untextured LoD3 models for analyzing solely influences of *geometric* building and façade properties, (b) textured meshes/LoD2 models (from Google Earth) as well as images from Google Street View for analyzing influences of *textural* information. Each of the three representation types are shown to the user in a way that at least two façades per building are visible. To avoid the assignment of the user being influenced by the building's environment, only the building itself appears; the environment of the building is not represented. Research on influences of the environment will be part of our future work.

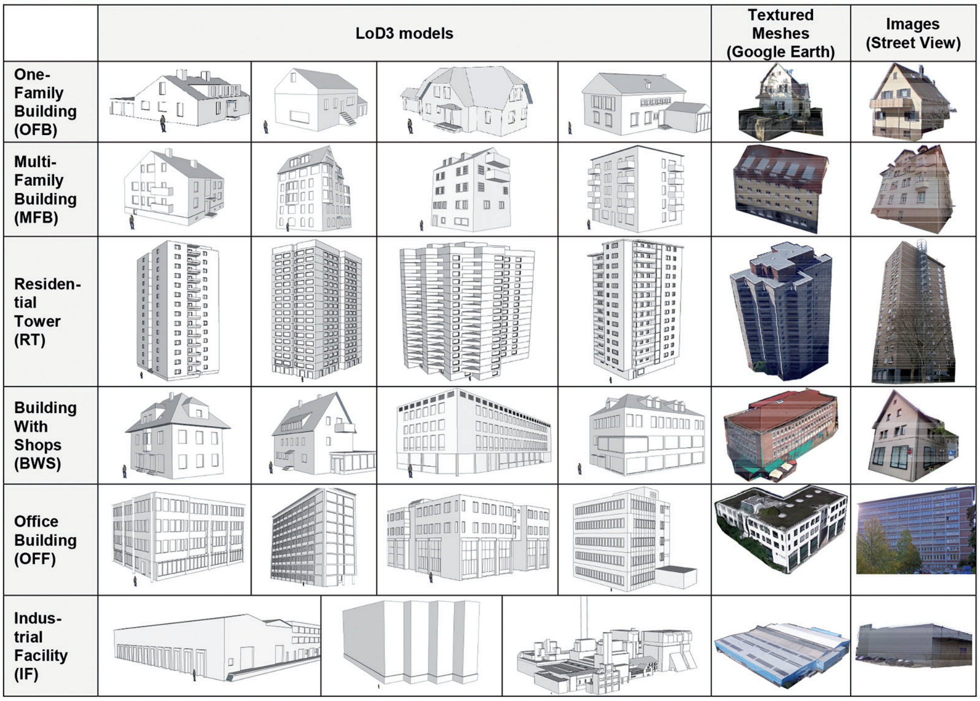

Fig. 1: Examples for building categories and representation types used in the study (Google Earth/Street View, ©2015 Google).

Within the study, users have to classify buildings into six characteristic building categories extracted from the ALKIS feature catalogue (ADV 2015):

- 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 are selected. For all these candidates LoD3 models have been modelled manually. For 60% of the buildings, additionally, textured meshes/LoD2 models from Google Earth and/or images from Google Street View are provided. Fig. 1 gives examples of the building categories and representation types presented to the user.

The user study is conducted as an online survey for the test person's convenience as well as faster evaluation reasons. At the beginning of the survey, some general information about the user is obtained, namely:

- Gender
- Age
- Graduation
- Subject of study
- Nationality
- Previous experiences in 3D virtual reality worlds (computer games, Google Earth, CAD modeling etc.)

Subsequently, the actual building category classification follows. All in all, 165 different building representations have to be classified by each participant. The representations are shown to the test person in random order. After the classification of each representation, users have to rate their level of certainty (reaching from 'Very Uncertain to 'Very Certain' in 5 selection options). Based on the self-

Fig. 2: Exemplary page of the study with a building model to be classified.

assessment for each classification, 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.

3.2 *Evaluation Metrics*

The actual reference category for each model is obtained by extracting the type of use from the digital city base map 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.

The 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, different appearance of ground floor compared to remaining floors, relative frequency of different roof types. The window-to-wall-surface ratio is given as the ratio of mean window surface area and the mean façade area (wall surface minus windows, doors etc.). Considering the property 'different appearance of the ground floor (GF) as compared to the remaining floors', 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. Considering the property 'different roof types', we discriminate between five different roof shapes: flat, saddle, hipped, monopitch and complex. Correspondingly, a roof complexity value ranging from 1 (simple) to 5 (complex) for each building category is computed as the weighted mean, with the weights being the occurring amount of each roof type within the class.

Based on these metrics, the discrepancy between ground truth and the user's perception is investigated. As a first step within this evaluation, the ground truth data is analysed. For each building presented to the user in the test, the above-stated features are determined. Since every building has been labelled into one of the 6 building categories presented in section 3.1, it is possible to calculate mean values of the features for each building category. These values can be considered representative for the respective category.

In a second step, the 6 building categories are set up again, however, 'as-perceived' this time. This means that for each category the entirety of all buildings classified into the respective class by all users is registered. Then again the mean values for each feature are computed, representing the 'as-perceived' or 'as-expected' features for each category. With this procedure, a comparison between the actual properties of a building class and the ones that were expected by the users is possible (see section 4.2.1).

4 Results and Application

This section is structured as follows: Overall results of the users' classification will be presented in section 4.1. Based on these results, concrete knowledge on the users' perception of building categories is derived in section 4.2. Finally, a first application of the obtained knowledge, namely perception-based abstraction, is presented in section 4.3.

4.1 *Classification Results of User Study*

In total, 96 test persons have participated in the user study. On average, the duration of the study was approximately 50 minutes. The participants' mean age is 24.8 years. The majority of the participants are students from Germany and abroad. In the following, we will first evaluate the classification results based on the entirety of all users (section 4.1.1). Afterwards, the results will be evaluated with respect to different groups of users (section 4.1.2).

4.1.1 Evaluation based on the entirety of all users

Tab. 1 depicts the confusion matrix for the building classification. Column headers 'GT' indicate ground truth.

The producer accuracy is given as the ratio of correctly classified buildings with regard to all ground truth buildings in this class. However, user accuracy is more interesting for this work – it is the fraction of correctly classified buildings with respect to all buildings classified to the current class. Commission errors correspond to buildings that were classified to a particular class, yet are actually belonging to another. Omission errors are buildings that actually belong to the ground truth class but were classified to a different category. The results can be seen in Tab. 2.

Obviously, One-Family Buildings and Industrial Facilities could be identified best with both over 90 percent user accuracy. Users have most difficulties with the classes Office Buildings, Building with Shops and Multi-Family Buildings which are indicated by user accuracies between 64.4% and 68.3%. Reasons for that will be further explained in section 4.2.

Besides the classification result, for each building the users should also rate their certainty for the 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 close-

Tab. 1: Confusion matrix for building classification (see section 3.1 for abbreviations of building categories).

	OFB GT	MFB GT	RT GT	BWS GT	OFF GT	IF GT	Sum Class
OFB	1483	59	1	69	1	2	1615
MFB	475	2462	137	460	62	8	3604
RT	6	377	2042	95	103	1	2624
BWS	23	222	87	1626	493	57	2508
OFF	18	237	513	265	1983	64	3080
IF	11	3	4	77	142	2172	2409
Sum GT	2016	3360	2784	2592	2784	2304	15840

Tab. 2: Classification metrics obtained from confusion matrix.

	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

ly to 'Certain'. This reflects the issue of the users who often not even know their current misinterpretation of the data. Even more: The user might feel certain in his 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.

4.1.2 Evaluation based on different groups of users

In the following, we will analyze 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), *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% foreign) as well as the users' *previous experience* with 3D virtual reality worlds (75% experience, 25% no experience). Thus, the user study is additionally evaluated with respect to the latter three properties. For this purpose, the same accuracy measures as in section 4.1.1 have been determined, this time, however, for the different user groups separately. Significance tests in form of Student's t-tests are carried out to search for significant differences in the classification results between those user groups.

Evaluation based on gender of users: 71% of the participants were male, 29% female. In order to examine whether there are gender-specific differences in the way of how humans perceive building categories, the results of both groups have been evaluated separately and compared to each other. The analysis shows no significant differences between male and female users.

Evaluation based on origin of users: To investigate influences of the user's origin on the classification results, an evaluation based on the user groups 'German' and 'foreign' has been performed. 38.5% of the users in the survey are from Germany, complementary 61.5% of the users have another nationality, distributed all over the world. Since all building models presented in the survey are located in Germany, and architectural construction for equal building types might vary throughout the world, this distinction seems eligible. However, tests on features in each building category did not reveal any significant difference between foreign and German users.

Evaluation based on users' previous experience: Further, the factor of self-assessment with regards to previous experiences in 3D virtual reality worlds is examined. 75% of the test persons stated that they have previous experience in this subject, whereas 25% stated they don't. However, the results for this subject are somewhat ambiguous, since experience in the topic of 3D virtual reality worlds could be interpreted quite widespread. Tests unveiled no significant difference between users with previous experience and novices.

As no significant differences in the classification results of the aforementioned user groups can be identified, all subsequent evaluations and interpretations in section 4.2 will be based on the entirety of all participants.

4.2 *Derivation of Knowledge on Building Perception*

Based on the findings described in section 4.1, we will now go a step further and try to derive coherences between the perceptibility of the building categories and several properties of the 3D representations. To find answers to questions such as 'Which representation type is for which building category the best?' or 'Which *geometric* building properties and structures are relevant for the perceptibility of a particular building category?', we proceed as follows: In section 4.2.1, we extract geometric dependencies, i.e., dependencies between the perceptibility of a building's category and the building's geometric properties. In section 4.2.2, the perceptibility with respect to different representation types is analyzed.

4.2.1 Perceptually relevant building structures

The goal is to derive geometric building properties and structures which are relevant or essential for the perceptibility of a specific building category. Following this goal, we first analyze the geometric properties of the building categories' representatives of our ground truth (see paragraph (a)). Afterwards, the same analysis is done for building categories as perceived by the users (see paragraph (b)).

(a) *Metrics of building categories reference*
The geometric building features introduced in section 3.2 are evaluated for each ground truth category (Tab. 3 (right part)). Based on that, it is tested whether the different building categories significantly differ in their geometric features. For that purpose, multiple significance tests are performed for each building feature's class mean. In the following, some significant characteristics for each building category within the ground truth are listed:

- ***One-Family Buildings*** have a significantly smaller footprint than all other categories besides *Buildings With Shops*. The total building height, the number of floors and the number of windows are smaller than in all other classes.
- For ***Multi-Family Buildings*** the total number of floors is significantly higher than for *One-Family Buildings* and *Industrial Facilities*, yet lower than for *Residential Towers* and *Office Buildings*. Accordingly the total number of windows is higher than for *One-Family Buildings* but lower than for *Residential Towers* and *Office Buildings*. *Multi-Family Buildings* only differ in few features from *Buildings With Shops*, hence the more important they are. The mean window surface is significantly smaller than for *Buildings With Shops*. Related thereto, a different arrangement, size, and shape of windows on ground floor and a different ground floor itself as compared to the remaining floors is significantly more important for *Buildings With Shops* than for *Multi-Family Buildings*.
- The most important feature of ***Residential Towers*** is the total number of floors, which is significantly higher than for all other building categories. Apart from *Multi-Family Buildings*, to which no significant difference is detected, the total amount of balconies is higher than in all other categories
- To distinguish ***Buildings With Shops*** from the rest, the most important features are different arrangement, size and shape of windows on ground floor as well as different ground floor itself in comparison to the remaining floors. These properties are significantly higher than in all other categories.
- Two features are salient for ***Office Buildings:*** The total amount of windows per façade, and the number of floors is significantly higher than for all other categories (except *Residential Towers*). Moreover, the mean entrance surface area is significantly higher than for *One-Family Buildings*, *Multi-Family Buildings* and *Residential Towers*. To distinguish *Office Buildings* from *Buildings With Shops*, 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 *Buildings With Shops*.

- For ***Industrial Facilities*** the footprint is the predominant feature because it is significantly higher than in all other categories. The window-to-wall-surface ratio is lower than for *Multi-Family Buildings*, *Residential Towers*, *Buildings With Shops* and *Office Buildings*.

(b) *Metrics of building categories as perceived/ expected by users*
As done before for the ground truth data, mean features are computed (Tab. 3 (left part)), this time based on the total amount of buildings all users classified into the respective class. To compare the ground truth data with the 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:

- For ***One-Family Buildings*** significant tests revealed, that there is no difference in perception and ground truth.
- For ***Multi-Family Buildings*** 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. Additionally, in the users' perception *Multi-Family Buildings* have a higher number of floors.
- To classify a building as ***Residential Tower***, for users, the number of floors can be less and the total height lower in comparison to ground truth. However, a single floor height is expected to be higher than for the ground truth.
- ***Buildings With Shops*** are considered to have a higher number of floors than in reality.
- For ***Office Buildings*** users are expecting a higher number of balconies.
- ***Industrial Facilities*** are expected to have more windows per façade and a bigger number of floors, too.

4.2.2 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. For 60% of the models at least two different representation types for the same building are available. 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. To determine whether the results actually differ from each other, again significance tests for the differences between geometric and textured representations results have been performed. 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 Fig. 1, all images are captured looking slightly upwards and thus resembling the human perspective. The viewing angle dependency on classification results is beyond the scope of this paper and will be addressed in further research.

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 somewhat more ambiguous categories like *Buildings With Shops*, *Office Buildings* and *Multi-Family Buildings* additional textural information improves the classification results.

4.3 *Application: Perception-Based Abstraction*

The knowledge derived in section 4.2.1 describes geometric 3D building properties and structures which are characteristic for a specific building category. A lot of applications where users have to move about in virtual 3D

Tab. 3: Geometric properties of the building categories as given in the ground truth (right part of the table), and as classified by the users in the study (left part of the table).

	As Classified						Ground Truth					
	IF	OFF	BWS	RT	MFB	OFB	IF	OFF	BWS	RT	MFB	OFB
Footprint (m^2)	912.6	480.00	432.80	328.95	252.92	148.69	10812.58	868.26	697.51	573.01	238.41	115.30
# Floors	5.0	6.3	5.7	6.9	5.8	3.1	2.5	5.9	3.7	15.5	4.0	2.1
Floor Height (m)	11.10	6.61	6.78	3.55	5.43	8.65	24.63	4.09	3.84	2.73	2.94	3.30
Total Height (m)	36.78	28.11	26.69	24.19	22.98	18.25	59.60	24.94	17.61	43.67	14.68	9.58
# Windows Per Façade	59.3	59.9	56.5	61.0	48.7	22.2	23.0	96.8	34.9	110.8	27.8	8.3
Ø Window Surface Area (m^2)	6.19	3.84	4.52	3.84	2.31	3.47	10.17	4.94	3.58	1.94	1.93	1.33
Window/Wall Surface Ratio (%)	61.7	52.5	53.6	56.9	32.5	25.9	10.1	127.9	52.2	29.1	30.0	16.1
# Entrances	2.8	2.3	2.3	1.4	2.2	1.6	5.3	1.4	1.6	1.2	1.4	1.7
Ø Entrance Surface Area (m^2)	13.48	8.97	9.86	8.39	7.35	4.99	13.33	7.93	24.90	3.53	2.99	4.13
# Balconies	1.5	2.1	2.0	3.0	2.6	0.9	0.0	0.1	1.1	9.5	2.2	0.2
Ø Balcony Surface Area (m^2)	3.88	3.34	3.30	4.36	3.82	1.80	0.00	3.42	10.08	4.24	2.06	1.04
Different Window Arrangement in GF	0.53	0.49	0.54	0.46	0.49	0.53	0.53	0.53	0.91	0.36	0.18	0.67
Different Window Sizes in GF	0.47	0.44	0.47	0.39	0.38	0.30	0.47	0.41	0.91	0.36	0.18	0.27
Different Window Shapes in GF	0.49	0.41	0.46	0.38	0.40	0.30	0.47	0.41	0.82	0.36	0.18	0.33
Different Ground Plan in GF	0.40	0.32	0.36	0.28	0.29	0.20	0.40	0.53	0.55	0.21	0.06	0.13
Roof Complexity	1.4	1.7	1.8	1.8	1.9	2.3	1.4	1.0	2.3	1.0	2.7	2.7

cities can benefit from this perceptual knowledge. Particularly when the virtual 3D city consists of abstracted, geometrically simplified buildings, e.g. when applications are visualized on small screens, it is even more important that the abstracted building representations still contain those geometric properties and structures which are essential for perceiving the correct category of the buildings. In the following, we will show how such perceptual knowledge can be embedded in a 3D abstraction process. Effects on the perceptibility of the buildings' categories will be demonstrated based on representative examples.

In a preprocessing step, information about perceptual relevance is attached to the respective 3D structures to provide semantically enriched building representations as input for the abstraction process. Based on Nan et al. (2011), we create different abstractions of buildings based on human perception. Nan et al. (2011) applied different Gestalt rules to drawings of façades which helped them to group drawing elements and to represent them by other elements. We extended this idea to the three-dimensional blocks formed by the façade elements and use it for abstracting given buildings. 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.

Fig. 3 (a1), (b1) and (c1) depict the original building models, respectively followed by two different results of the abstraction process. For the first abstraction, parameters have been chosen based on features that are important for the user to correctly classify a building. The second abstraction is completely free, meaning that no restrictions were made during the abstraction. Fig. 3 (a) depicts a model belonging to the class of *Building With Shops*. The first abstraction incorporates the properties learned to be important for *Building With Shops* as mentioned in section 4.2.1, paragraph (a). 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. This merged window front destroys the building's original property of having significantly bigger windows in the lower floors than in the remaining floors, which was detected to be an important feature of *Buildings With Shops*, though.

In Fig. 3 (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* (Tab. 3). The important feature 'balcony' is maintained in the first abstraction, the second abstraction however drops it. This way, model (b2) retains the appearance of a residential building, whereas model (b3) is more neutral and, thus, could also be interpreted as an *Office Building*.

Fig. 3 (c) shows the example of an *Office Building*. The abstracted model (c2) keeps the characteristic structure of the ground floor but merges windows in the upper floors, thus still closely resembling the original. Model (c3) though merges windows and entrances in the ground floor. As a consequence, the model might rather be perceived as *Building With Shops* than *Office Building*.

5 Conclusions and Outlook

With the aim of deriving 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. Within the study, the users were asked to classify consecutively presented single building representations into the categories *One-Family Building*, *Multi-Family Building*, *Residential Tower*, *Building With Shops*, *Office Building* and *Industrial Facility*. During

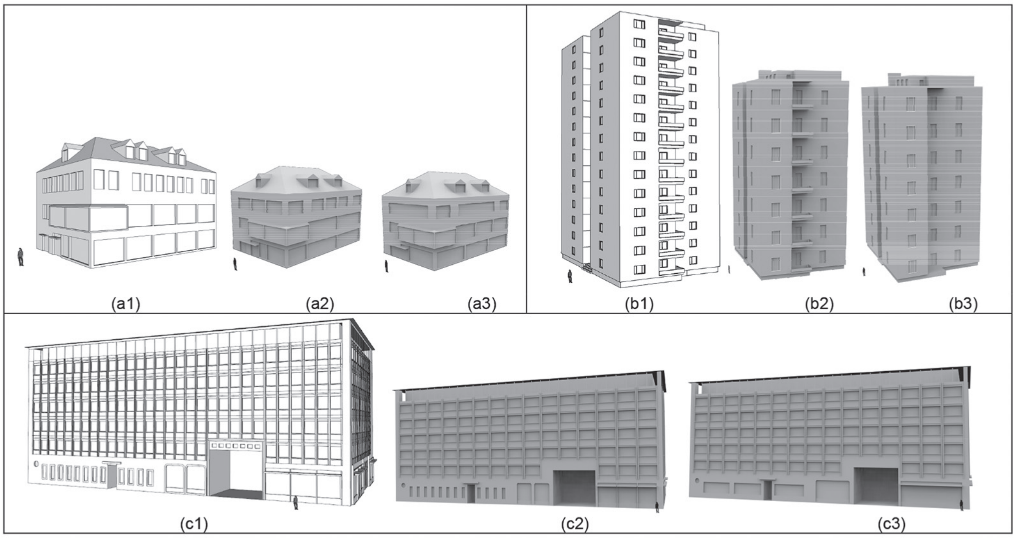

Fig. 3: Application of conclusions drawn from the survey. For x∈{a,b,c}: (x1) original building model, (x2) abstraction based on features important for the user to classify into the respective correct category, (x3) 'free' abstraction without restrictions.

the whole classification process, the users additionally had to rate their level of certainty. The representations shown to the users were untextured LoD3 models, textured meshes/ LoD2 models from Google Earth, and images extracted from Google Street View.

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: The overall classification accuracies for textured meshes/LoD2 models from Google Earth and images from Google Street View are 75.4% and 79.3% and, thus, significantly higher than the classification accuracy of untextured LoD3 models, which lies at 69.2%. Particularly for buildings that are belonging to somewhat more ambiguous categories like *Buildings With Shops*, *Office Buildings* and *Multi-Family Buildings* additional textural information improves the classification results. However, for building categories that are easily separable from the rest like *One-Family Buildings* and *Industrial Facilities*, a geometric representation in form of a LoD3 model is sufficient in the majority of cases.

The classification accuracy of LoD3 models mainly depends on whether the building models show properties that have been detected as perceptually relevant for the respective building category. Examples for such perceptually relevant geometries and structures are the occurrence of balconies for *Residential Towers*, the different appearance of ground floor and remaining floors for *Buildings With Shops*, or the high windows-to-wall-surface ratio for *Office Buildings*. As these properties are not inherent in all representatives of the categories mentioned, users sometimes experience difficulties to distinguish between *Buildings With Shops*, *Multi-Family* and *Office Buildings*. Moreover, 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. The knowledge gathered in the investigation of ground truth features and the significant features as-perceived or expected by users can then be used to generate virtual 3D models that support and improve the correct perception of building categories.

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. Characteristic properties of building structures that turned out to be essential for the perceptibility of a certain building category are maintained during the abstraction process whereas structures that are unimportant or even obstructive are simplified to a much higher extent or even totally neglected. Doing so, the perceptibility of the building category can be preserved even in abstracted building representations.

In our future work, we plan to further extend our findings about the human ability to understand building categories based on user studies which will be implemented in an interactive 3D visualization software. Having considered buildings so far on an individual basis only, we will investigate how the human perception of a building's category is influenced by the building's environment. Further experiments will be set up, for example, to retrieve information about the impact of different building representations on the users' way of navigating through virtual 3D environments. The perceptual knowledge gained from those analyses will be embedded into a framework with which it will be possible to not only *maintain* perceptually relevant geometric properties and structures as it was the case in the perception-guided abstraction but it will additionally allow to *modify*, *emphasize*, *add* or *remove* perceptually relevant structures in a targeted manner in order to automatically generate models that can be classified more easily into the respective correct building category.

Acknowledgements

We would like to thank the German Research Foundation (DFG) for financial support within the projects D01, A04 and task force TF3 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

Adabala, N., 2009: Building representation in oblique-view maps of modern urban areas. – Cartographic Journal **46:** 104–114.

AgarwalL, S., Snavely, N., Simon, I., Seitz, S.M. & Szeliski, R., 2009: Building rome in a day. – IEEE 12th international conference on computer vision: 72–79.

AdV, 2015: Dokumentation zur Modellierung der Geoinformationen des amtlichen Vermessungswesens. – Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland, http://www.adv-online.de/AAA-Modell/binarywriterservlet?imgUid=3c860f61-34ab-4a41-52cf-b581072e13d6&uBasVariant=11111111-1111-1111-1111-111111111111 (25.7.2016).

Cao, F., Delon, J., Desolneux, A., Musé, P. & Sur, F., 2007: A unified framework for detecting groups and application to shape recognition. – Journal of Mathematical Imaging and Vision **27:** 91–117.

Desolneux, A., Moisan, L. & Morel, J.-M., 2004: Seeing, Thinking, Knowing. – Carsetti, A. (ed.): Kluwer Academic Publishers: 71–101.

Engel, J., Schöps, T. & Cremers, D., 2014: LSD-SLAM: Large-scale direct monocular SLAM – European Conference on Computer Vision, Springer International Publishing: 834–849.

Fritsch, D., Khosravani, A., Cefalu, A. & Wenzel, K., 2011: Multi-sensors and multiray reconstruction for digital preservation. – Fritsch, D. (ed.): Photogrammetric Week, Wichmann: 305–323.

Glander, T. & Döllner, J., 2009: Abstract representations for interactive visualization of virtual 3D city models. – Computers, Environment and Urban Systems **33:** 375–387.

Gröger, G. & Plümer, L., 2012: CityGML – Interoperable semantic 3D city models. – ISPRS Journal of Photogrammetry and Remote Sensing **71:** 12–33.

Haala, N., 2013: The landscape of dense image matching algorithms. – Fritsch, D. (ed.): Photogrammetric Week. – Wichmann: 271–284.

Hirschmüller, H., 2008: Stereo processing by semiglobal matching and mutual information. – IEEE Transactions on pattern analysis and machine intelligence **30** (2): 328–341.

Kolbe, T.H., Gröger, G. & Plümer, L., 2005: CityGML – interoperable access to 3D city models. – Oosterom, Zlatanowa & Fendel (eds.): International Symposium on Geoinformation for Disaster Management: 883–899, Springer.

Kubovy, M. & Van den Berg, M., 2008: The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. – Psychological Review **115** (1): 131–154.

Li, Z., Yang, H., Ai, T. & Chen, J., 2004: Automated building generalization based on urban morphology and Gestalt theory. – International Journal of Geographical Information Science **18:** 513–534.

Mayer, H., Bartelsen, J., Hirschmüller, H. & Kuhn, A., 2012: Dense 3D reconstruction from wide baseline image sets. – International Conference on Theoretical Foundations of Computer Vision: 285–304, Springer-Verlag, Berlin, Heidelberg.

Michaelsen, E., Iwaszczuk, D., Hoegner, L., Sirmacek, B. & Stilla, U., 2012: Gestalt grouping on facade textures from IR image sequences: comparing different production systems. – The ISPRS International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences: 303–308.

Michaelsen, E. & Yashina, V., 2013: Simple Gestalt algebra. – International Workshop on Image Mining, Theory and Applications: 3–13.

Nan, L., Sharf, A., Ke, X., Deussen, O., Cohen-or, D. & Chen, B., 2011: Conjoining Gestalt rules for abstraction of architectural drawings. – ACM Transactions on Graphics **30** (6): 185:1–185:10.

Pasewaldt, S., Semmo, A., Trapp, M. & Döllner, J., 2014: Multi-perspective 3D panoramas. – International Journal of Geographical Information Science **24** (10): 1–22.

Sun, X., Yang, B., Attene, M., Li, Q. & Jiang, S., 2011: Automated abstraction of building models for 3D navigation on mobile devices. – 19th International Conference on Geoinformatics, Shanghai, China, 24–26 June, 6 p.

Addresses of the Authors:

Dipl.-Ing. Patrick Tutzauer, Dr.-Ing. Susanne Becker and Prof. Dr.-Ing. Dieter Fritsch, Universität Stuttgart, Institut für Photogrammetrie, Geschwister-Scholl-Str. 24d, 70174 Stuttgart, Tel.: +49-711-685-84093, Fax: +49-711-685-83297, e-mail: {patrick.tutzauer}{susanne.becker}{dieter.fritsch}@ifp.uni-stuttgart.de

M. Sc. Till Niese and Prof. Dr. Oliver Deussen, Universität Konstanz, Fachbereich Informatik und Informationswissenschaft, Fach 698, 78457 Konstanz, Tel.: +49-7531-88-4233, Fax: +49-7531-88-4715, e-mail: {till.niese}{oliver.deussen}@uni-konstanz.de

Manuskript eingereicht: July 2016
Angenommen: September 2016