

# Digital Material Appearance: The Curse of Tera-Bytes

by Michal Haindl, Jiří Filip and Radomír Vávra

*Real surface material visual appearance is a highly complex physical phenomenon which intricately depends on incident and reflected spherical angles, time, light spectrum and other physical variables. The best current measurable representation of a material appearance requires tens of thousands of images using a sophisticated high precision automatic measuring device. This results in a huge amount of data which can easily reach tens of tera-bytes for a single measured material. Nevertheless, these data have insufficient spatial extent for any real virtual reality applications and have to be further enlarged using advanced modelling techniques. In order to apply such expensive and massive measurements to a car interior design, for instance, we would need at least 20 such demanding material measurements.*

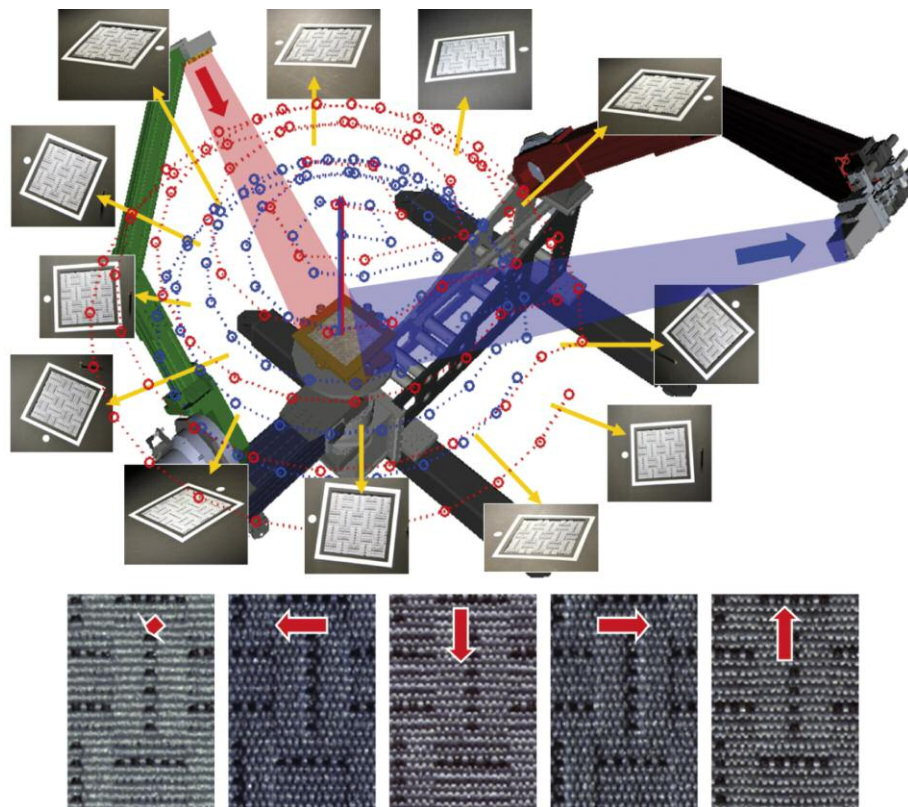
Within the Pattern Recognition department of UTIA, we have built a high precision robotic gonioreflectometer [2] (see Figure 1). The setup consists of independently controlled arms with camera and light. Its parameters such as angular precision (to 0.03 degrees), spatial resolution (1000 DPI), and selective spatial measurement qualify this gonioreflectometer as a state-of-the-art device. The typical resolution of an area of interest is around 2000 x 2000 pixels, each of which is represented by at least 16-bit floating point values to

achieve reasonable representation of high-dynamic-range visual information. The memory requirements for storage of a single material sample amount to 360 giga-bytes per colour channel. Three RGB colour channels require about one tera-byte. More precise spectral measurements with a moderate visible spectrum (400-700nm) sampling further increase the amount of data to five tera-bytes or more. Such measurements (illustrated in Figure 2) allow us to capture the very fine meso-structure of individual entities comprising the material.

Such enormous volumes of visual data inevitably require state-of-the-art solutions for storage, compression, modelling, visualization, and quality verification. Storage technology is still the weak link, lagging behind recent developments in data sensing technologies. Our solution is a compromised combination of fast but overpriced disk array and slow but cheap tape storage. The compression and modelling steps are integrated due to our visual data representation based on several novel multidimensional probabilistic models. Some of these models (using either a set of underlying Markov random fields or probabilistic mixtures) are described in ERCIM News 81 [1]. These models [2] allow unlimited seamless material texture enlargement, texture restoration, huge unbeatable appearance data compression (up to 1:1000 000) and even editing or creation of novel material appearance data. They require neither storing of original measurements nor any pixel-wise parametric representation.

A further problem is that of physically correct material visualization, because there are no professional systems which allow rendering of such complex data. Therefore, we were forced to develop the novel Blender plugin for the purpose of realistic material appearance model mapping and rendering.

*Figure 1: Appearance measurement principle. The total number of images is equal to every possible combination of red dots (camera directions) and blue dots (light directions). Below is an example of how the visual appearance of fabric is dependent upon illumination direction. Fabric is illuminated from above, left, top, right and below, respectively.*



Blender is a free open source 3D graphics application for creation 3D models, visualizations and animations and is available for all major operating systems under the GNU General Public License. Visual quality verification is another difficult unsolved problem which we tackle using applied psychophysically validated techniques.

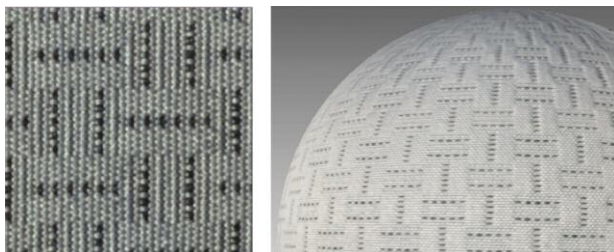


Figure 2: Detail of measured fabric material (left) and its large-scale mapping on a sphere (right).

These precise visual measurements are crucial for a better understanding of human visual perception of real-world materials. This is the key challenge not only for efficient and simultaneously convincing representations of visual information but also for further progress in visual scene analysis. Due to their tricky and expensive measurements only a small amount of data of low quality are available so far. Therefore, we provide benchmark material measurements for image and computer graphics research purposes on our web server listed below.

**Link:**

<http://btf.utia.cas.cz/>

**References:**

- [1] Haindl M., Filip J., Hatka M.: Realistic Material Appearance Modelling . ERCIM News (2010), 81, pp.13-14
- [2] Haindl, M., Filip, J. Advanced textural representation of materials appearance, Proc. SIGGRAPH Asia '11 (Courses), pp. 1:1-1:84, ACM.

**Please contact:**

Michal Haindl  
 CRCIM (UTIA), Czech Republic  
 Tel: +420 266052350  
 E-mail: [haindl@utia.cas.cz](mailto:haindl@utia.cas.cz)

## (No)SQL Platform for Scalable Semantic Processing of Fast Growing Document Repositories

by Dominik Ślęzak, Krzysztof Stencel and Son Nguyen

*Large volumes of scientific data require specific storage and indexing solutions to maximize the effectiveness of searches. Semantic indexing algorithms use domain knowledge to group and rank sets of objects of interest, such as publications, scientists, institutes, and scientific concepts. Their implementation is often based on massively parallel solutions employing NoSQL platforms, although different types of semantic processing operations should be scaled with respect to the growing volumes of scientific information using different database methodologies [1].*

SONCA (Search based on ONtologies and Compound Analytics) platform is developed at the Faculty of Mathematics, Informatics and Mechanics of the University of Warsaw. It is part of the project ‘Interdisciplinary System for Interactive Scientific and Scientific-Technical Information’ ([www.synat.pl](http://www.synat.pl)). SONCA is an application based on a hybrid database framework, wherein scientific articles are stored and processed in various forms. SONCA is expected to provide interfaces for intelligent algorithms identifying relations between various types of objects [2]. It extends typical functionality of scientific search engines by more accurate identification of relevant documents and more advanced synthesis of information. To achieve this, concurrent processing of documents needs to be coupled with ability to produce collections of new objects using queries specific for analytic database technologies.

SONCA’s architecture comprises four layers (Figure 1). User Interface receives requests in a domain-specific language, rewrites them into the chains of (No)SQL statements executed iteratively against the Semantic Indices and – in some cases – the contents of SoncaDB, and prepares answers in a form of relational, graph or XML structures that can be passed to external reporting and visualization modules.

The Semantic Indices are periodically recomputed by Analytic Algorithms basing on continuously growing contents of SoncaDB. SoncaDB stores articles (and other pieces of information) acquired from external sources in two formats: XML structures extracted for each single document using structural OCR, and subsets of tuples corresponding to documents’ properties, parts, single words and relationships to other objects, populated across data tables in a relational database.

The roles of Analytic Algorithms and the relational subpart of SoncaDB are two examples of SONCA’s innovation. Additional storage of full information about articles in a tabular form gives developers of Analytic Algorithms a choice