Editor: **Steffen Staab,** University of Koblenz-Landau, staab@uni-koblenz.de

3D Media and the Semantic Web

Michela Spagnuolo and Bianca Falcidieno, *National Research Council of Italy*

3D content is widely recognized as the next wave of digital media. The success of 3D communities and mapping applications (for example, Second Life and GoogleEarth) and the decreasing costs of producing 3D environments are leading analysts to predict a dramatic shift in how people see and navigate the Internet. Greg Sterling, founder of the research firm Sterling Market Intelligence, suggests that "the Internet could very well be on its way to shifting from a text-based environment to a visually oriented 3D world."[1]

We can liken 3D content's impact on Internet-based applications with the impact images have had, with a number of distinctions. 3D media offers greater potential for interactivity because users can observe and manipulate them from different viewpoints. However, representing a complex shape is not trivial owing to the volume of data involved, the variety of representation models, and the complexity and heterogeneity of meaning and semantics that 3D content can reveal (see Figure 1).

The ease of producing and collecting digital data has caused a gradual paradigm shift in various applied and scientific fields: from physical prototypes and production to virtual prototypes and simulation. This shift has had an enormous impact in domains where 3D media are essential knowledge carriers and represent a huge economic factor. Such domains include design and manufacturing, serious gaming and simulation, cultural heritage and archaeology, medical applications, bioinformatics, and pharmaceutical science.

In domains such as these, where workflows for 3D content creation and analysis will soon become completely digital and benefit from a collaborative environment, the Semantic Web could prove an extremely powerful supporting mechanism. But the computer graphics community faces challenges to fill the semantic gap for 3D media and equip it with tools that allow interoperability through formal languages and semantics.

3D: What's That?

3D models are digital representations of objects that can be processed by computer applications. The objects themselves may exist either in the physical or virtual world. A virtual modeling system can generate digital representations, or you can create them by acquiring dense data from existing physical objects—for instance, by photogrammetry, computerized tomography, laser scanning, or other digitization technologies.

Digital representations of geometric models can represent not only solid objects but also any other phenomenon realized in the 3D space—for instance, molecular surfaces, electromagnetic fields, and environmental data. All shapes can be described by their geometry (the object's spatial extent), structures (object features and part-whole decomposition), attributes (colors, textures, and names attached to an object, its parts, and its features), semantics (meaning and purpose), and interaction with time (that is, history, shape morphing, animation, and video).

Traditional 2D media such as images or videos only partially represent an object's shape, owing to projection and occlusion. But when we deal with 3D media, we typically assume a complete specification of the object's shape. This representation allows for higher degrees of interaction than with 2D media: 3D models can be turned, scaled, viewed in any direction, manipulated, and used to compose new scenes about physical or simulated environments where users can experience an immersive presence.

The synthesis, analysis, processing, and visualization of 3D media are traditional fields of computer graphics expertise based on a wide spectrum of fun-

1541-1672/09/$25.00 © 2009 IEEE
Published by the IEEE Computer Society

damental domains, comprising computational geometry, algebraic geometry, discrete topology, differential geometry, numerical analysis, and linear algebra. In the last decade, computer graphics has matured to the point where the fundamentals of modeling, visualization, and streaming of static and dynamic 3D shapes are well understood.

Different geometric models can represent the same object, either as a composition of simpler volumes or as a tessellation of its enclosing boundary. Nowadays, geometric-model variants range from point set to subdivision surfaces and from implicit to skeletal representations, but the triangle mesh is probably the most common. Triangle meshes represent objects through a triangular network among points scattered on the shape's boundary surface or inside its enclosed volume. Obviously, each representation schema corresponds to an abstraction of the object shape that answers specific application requirements. For this reason, a representation schema isn't interchangeable in a straightforward manner. Triangle meshes are becoming the standard in many applied fields, especially for Web-based applications, because their irregular structure supports modeling at various resolutions and progressive transmissions of data (see Figure 2).

From Web Applications to Science and Industry

Until the last decade, experienced professionals were the main producers of 3D content. Today, regular users can easily create new content in virtual 3D environments. Soon, users will be able to acquire physical assets or convert 2D data into 3D data to enhance their virtual-world experience.

Despite the questionable hype of 3D social networking on sites such as Second Life, a number of companies are producing 3D city models and building Web-based applications to create scenarios for 3D data use. Besides the realm of entertainment, virtual environments have great potential for simulation and serious gaming. They can serve as the background to a new generation of graphical interfaces for a variety of collaborative working setups:

- to build virtual laboratories where scientists can work with and share digital 3D models of physical or theoretical objects,
- to train professionals in complex maintenance scenarios,
- to remotely access virtual museums in distant geographical locations, and
- to generally incorporate nonverbal communication into same-time, different-place interactions.

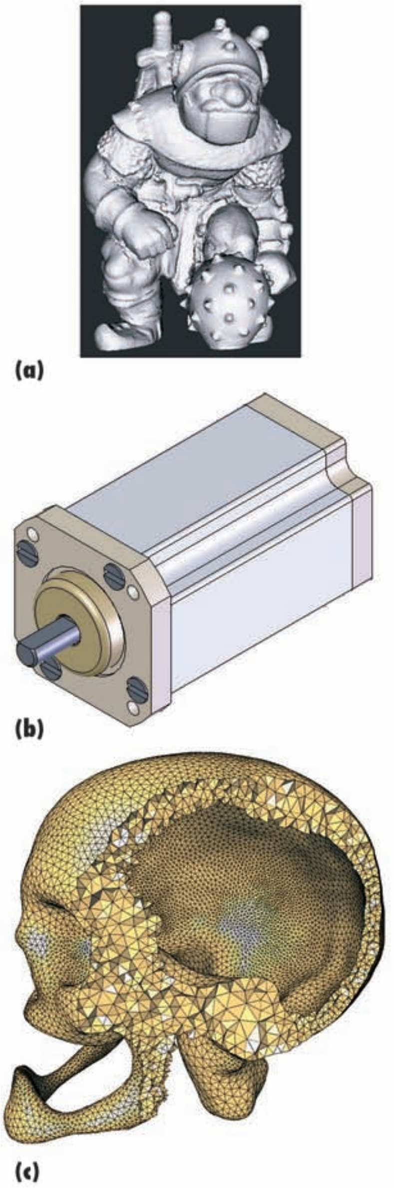

Figure 1. 3D models with different representation models: (a) triangle mesh, (b) B-rep, and (c) volumetric mesh. Note how the type and scale of the object features vary. Models courtesy of the AIM@SHAPE Repository.

Figure 2. A triangle mesh at different resolution levels. The same shape is represented with different vertex and triangle counts, according to the level of detail required by the application or by streaming requirements.

With many applications relying on 3D and networked environments, we can see the 3D Internet emerging. The hardware technologies are ready, 3D modeling methodologies are mature, and bandwidth requirements remain high but will likely be met soon. In five years, the 3D Internet will be as important for work as the Web is today. This rapid technological evolution raises new challenges in terms of answering the emerging needs of the variegated community of professionals and novice users who will soon face the problems of administering, structuring, and accessing the amounts of information carried by 3D content. We believe that an early adoption of Semantic Web techniques is the key to shaping the growth of next-generation 3D- and knowledge-intensive application domains that will rely heavily on the Internet as the collaboration framework.

Embedding 3D into the Semantic Web

The Semantic Web is basically a set of tools that supports interoperability between humans and machines through formal syntaxes, ontologies, and inferencing rules.[2] The adoption of Semantic Web methods could bring a number of concrete contributions to the development of the 3D Internet and its applications. There are obvious advantages for scientific and industrial fields dealing with 3D data. Internet-based services—on the Web or local networks—for automatic content and knowledge resource location would enhance data exploration and scientific hypothesis validation, analysis, and comparison.[3] In unspecialized scenarios, searching by content, collaborative tagging, and 3D reconstructions from tagged images or videos could be coupled with easy-to-use Web-enabled modeling tools to create new applications that cross the boundaries of virtual worlds and bring virtual 3D into the physical world. The impressive achievements of 3D printer technologies enable *fabbing*—creating a physical instance of a digital object by printing it out in 3D. The techniques are now so mature and versatile that you can print not only mockups but even regular end products in a variety of materials.[4]

We have plenty of tools for visualizing, streaming, and interacting with 3D objects, even in unspecialized Web contexts. However, tools for coding, extracting, sharing, and retrieving the semantic content of 3D media are still far from satisfactory. This hinders—in the short term—the application of existing Semantic Web techniques to share and retrieve resources as well as extract new knowledge using more sophisticated inference mechanisms. The computer graphics community is now well aware of the potential benefits of a shift toward semantics-oriented modeling. Automatic classification of 3D databases, 3D content annotation, and content-based retrieval have introduced new areas of research that represent some of the key topics in computer graphics and vision research.

Tools for coding, extracting, sharing, and retrieving the semantic content of 3D media are still far from satisfactory.

Documenting 3D Media for Sharing

Knowledge technologies, even basic ones, didn't influence the development of 3D modeling approaches and sharing practices until few years ago.

The Stanford 3D Scanning Repository (http://graphics.stanford.edu/data/3Dscanrep), the best-known example in computer graphics, dates back to the late '90s. The "Stanford Bunny," a digital model of a clay rabbit, appeared in countless journal papers and became a de facto standard test model in computer graphics. The Stanford repository, however, is simply a Web page with links to data sets, models, and software tools to visualize or convert model formats. The Web page describes models with textual information that's not organized as metadata. In practice, no services are enabled on the repository.

The first repository to exhibit some structured organization and basic services was the Princeton Shape Benchmark, built to provide 3D models and software tools for evaluating shape-based retrieval and analysis algorithms (http://shape.cs.princeton.edu/benchmark). The repository contains models collected from the Web and manually classified in a variety of functional and shape categories, such as animals, airplanes, boats, and furniture. Each model is linked to an information file containing data about the original Web site the model comes from and some geometric metadata, such as the number of polygons and connected components.

The Digital Shape Workbench (DSW, http://dsw.aimatshape.net), developed in the *A*dvanced and *I*nnovative *M*odels *A*nd *T*ools for the Development of *S*emantic-Based systems for *H*andling, *A*cquiring, and *P*rocessing Knowledge *E*mbedded in Multidimensional Digital Objects (AIM@SHAPE) Network of Excellence project, represents a unique infrastructure prototype for storing scientific resources and supporting research activities. The infrastructure is knowledge based to take into account taxonomies of geometric representations and ontologies describing workflows of 3D modeling and processing operations, while also implementing various search mechanisms.

The AIM@SHAPE model repository stores more than 700 high-quality models, each characterized by a large set of metadata spanning from simple Dublin Core to more specific geometry-oriented metadata, structured according to the Common Shape Ontology. The ontology is expressed in OWL, while RacerPro provides reasoning on the knowledge base. The ontologies were mainly intended for computer graphics researchers, so they primarily concern properties of the representation, topological characteristics, model size, and so on. For some classes, the geometric metadata can be extracted automatically during the upload.

The model repository also provides services for downloading models at user-required resolution levels and for visualizing the content before downloading it—important features when dealing with large data sets.

Similarly, the AIM@SHAPE tools repository stores 70 tools for processing 3D models documented with metadata derived from a task-driven ontology. The repositories gained a reputation in the graphics community owing to the detail of the metadata used to describe the content and the upload and download services provided.

As a proof of concept, the DSW demonstrated that exposing the semantics associated with 3D models—at least at a geometric level—facilitates sharing and reuse of the data and tools used to process them. In other words, applying Semantic Web technologies to the integration and correlation of distributed and heterogeneous resources about 3D content—models, tools, workflows—offers enormous potential for accelerating the discovery of new knowledge by reusing existing resources.

Understanding the Meaning of 3D Media

Formalizing and structuring metadata about the representation and format of 3D data isn't enough to answer the needs of a wider community of 3D users. A semantic description of 3D objects is commonly meant to describe the content with meaningful terms from some domain of knowledge. For example, a given model can be described as a table made of four cylindrical legs and an oval top. Current 3D modeling systems handle the geometric representation of digital shapes but not their semantics—that is, the meaning or functionality in a given context. Consequently, geometric representations don't give explicit information about the content's semantics, which you can grasp only by viewing the object. If we want machines to understand the content of digital 3D media, we need tools to automatically classify objects in semantic classes, extract salient features, and segment the representation into meaningful parts. Basic elements in a 3D representation—points, triangles, and lines—don't know which semantic unit they belong to.[5]

Basic elements in a 3D representation—points, triangles, and lines—don't know which semantic unit they belong to.

Much of computer graphics deals with shape analysis and segmentation. Apart from a few specific algorithms, semantic properties are taken into account, mainly implicitly, at the level of cognitive theories used to describe the segmentation approach—namely, part-based decomposition or the minima rule. Part-based decomposition methods build on Irving Biederman's theory of perception, which characterizes an object as a compound of primitive basic parts—for example, cylinders or cubes.[6] The minima rule suggests that we perceive relevant parts by focusing on lines of concave discontinuity of the tangent plane.[7] Ariel Shamir provides a recent survey of segmentation methods.[8]

So far, most methods developed for shape analysis and segmentation don't directly output semantically relevant explicit shape descriptions. Instead, they provide rich characterizations of the geometric and structural object boundary properties. For CAD models, it's easier to devise a feature extraction method, because engineering features have a rather well-formalized characterization. Conversely, feature recognition for free-form shapes is challenging, especially when the semantics underlying the features are related to an intrinsically vague context, as with most natural shapes.[9] Think, for example, of the human body's semantics: neck, legs, thigh, elbow, and many other terms that identify relevant body parts. Refer to portions of the body shape that can't be precisely coded or identified by a mathematical formulation and whose boundaries can't be drawn precisely. Also, some body features consist of other features; a leg is defined by the shin, the calf, and the thigh. Its articulation depends on the knee, whereas the ankle and hip connect it to the body. To our knowledge, only one paper has addressed feature recognition using an underlying feature ontology in a very focused domain.[10] Although it is rather a toy example, the paper proposes that we could achieve good results only by coupling geometric segmentation and structuring methods with ontologies and then reasoning on the segmentation output.

Annotation and Markup of 3D Media

Shape segmentation and feature extraction are the computational tools needed to select the portion of interest in a model. Selecting regions of interest in the manual annotation of 2D media is rather simple in terms of user-interface: dragging a selection box or lasso tool over an image achieves the necessary functionality. The same doesn't hold for 3D media, or at least it's complicated by the data's nature: parts might be out of reach for mouse interaction, and bounding a part can be rather complex.

Shape segmentation is therefore complemented by shape annotation, which is typically the means to document content with contextual knowledge, either manually or automatically. Because 3D annotation is a relatively new topic, few prototypes exist.[11,12]

The two most important aspects of an annotation framework involve what is annotated and how it is annotated. For the first aspect, it's important to understand whether the target is the whole object or its constituent parts. If the target is constituent parts, the annotation must be combined and output using shape segmentation. This case is promising and flexible, and supports content-based retrieval using textual queries that take into account annotations about objects and their parts. For the second aspect, users may want to annotate 3D objects with free or ontology-driven tags. While the first approach is more suitable to applications inspired to social networking principles, the second is more suitable for collaborative working environments where annotations serve the purpose to enrich geometric data with additional knowledge. Knowledge about 3D content may be:

- Knowledge related to the geometry of 3D media. The object's geometry remains the same in every context. Knowledge concepts at this level could be shared by all 3D applications.
- Knowledge related to the meaning of the object represented in the 3D media. 3D media can represent objects that belong to a category of shapes, either in broad, unrestricted domains (for example, chairs and tables in furniture) or narrow, specific domains (for example, T-slots and pockets in mechanical engineering). Feature concepts should allow for different granularity and specialization levels.
- Knowledge related to the application domain in which 3D media are manipulated. The application domain, especially in scientific and industrial scenarios, defines how the 3D shape should be represented, processed, and interpreted. The knowledge of domain experts plays a big role here in manipulating the digital model and devising ad hoc solutions for specific domains.

The two most important aspects of an annotation framework involve what is annotated and how it is annotated.

Two outstanding issues exist here. The first is how to structure the annotation process's output. The second is how to deal with the transient nature of geometric primitives with respect to persistency of meaning, as Sven Havemann and Dieter W. Fellner described.[4]

Concerning the first issue, there's no clear approach to linking annotations to 3D media. Current standards for expressing geometric data, such as X3D, allow the possibility of describing compound scenes as assemblies of simpler ones and describing behaviors and interactions, but this is mainly considered a standard to code information needed for interactive applications rather than for a 3D Semantic Web. Traditional geometric representations, in various types, could be enhanced with tags, annotations, and even hyperlinks to other 3D models, making them evolve toward what Havemann and Fellner call generalized 3D documents.[4] Marco Attene and his colleagues' ShapeAnnotator keeps the geometry and annotation in two distinct files; the annotation produces a set of instances that, together with the domain ontology, form a knowledge base.[11] Each instance is related to one part of the model, and it's defined by its URI, its type (the class the feature belongs to) and other attribute values and relations. In its current version, the ShapeAnnotator saves the geometric representation, augmented with information about the segmentation, into a single file, and saves the instances as a separate OWL file that imports the domain ontology.

This is a partial solution. The problem of defining a stable 3D markup remains.

Concerning the geometric primitives issue, annotations or tags attached to parts of 3D models should survive changes in the geometric representation. This task isn't trivial, even if we consider just one representation type, such as triangle meshes. Think of part-based annotation of a 3D model representing a statue or a complex artifact. For visualization, we need to simplify the models; that is, we need to remove a number of vertices and triangles. What happens to the annotations? How do we keep them consistent across resolution changes? The problem gets even more complicated if we think of completely changing the representation type—for instance, switching from triangle to quadrilateral meshes. The statue's shape, together with its relevant features, remains the same, and the annotations should follow the scale changes accordingly and smoothly.

Content-Based Retrieval of 3D Media

In the current panorama of 3D search engines, there's no satisfactory way to search for a semantically relevant feature—a house or a tree—in a 3D scene or 3D models repository, unless the 3D models have been manually annotated with keywords, which reduces the problem to a text search.

Researchers are exploring content-based search to overcome this problem by letting users search for 3D content resembling a sketched query or by using a query-by-example. Nothing has been developed into a full-scale search engine, and the results such systems produce are far from the results obtained by text searches. From a high-level perspective, the main components of a content-based retrieval system for 3D or 2D visual media are similar:

- a feature extraction module and indexing system that usually works offline,
- a query formulation and result delivery module, and
- an online matching module that extracts the most relevant items from a collection according to metrics defined on the feature vector space.

The main differences between the 2D and 3D context arise during feature extraction, and unfortunately most 2D methods don't generalize to 3D because of the context's different nature.

Assessing similarity among 3D shapes is a complex, challenging research topic, and researchers have only recently addressed the computational aspects of 3D shape retrieval and matching.[13,14] The methods developed so far span from coarse filters suited to browse large 3D repositories on the Web, to domain-specific approaches for assessing similarity of part models containing semantic and structural information. Most proposed methods focus on the geometry of shapes, in the sense of considering spatial distribution or extent in the 3D space. Nevertheless, consensus exists that people recognize and mentally code an object's shapes in terms of relevant parts and their spatial configuration or structure. The use of structural descriptions supports reasoning on shape similarity at a local or partial level.

> People recognize and mentally code an object's shapes in terms of relevant parts and their spatial configuration or structure.

To be really helpful in the Semantic Web sense, next-generation 3D search engines should be able to integrate content-based, geometry-driven criteria with concept-based, semantically driven criteria. Such engines should allow queries such as "find the 3D models in the repository that represent a vase with handles, and whose handles are globally similar in shape to a given query model." In the example, "vase" and "handle" could refer to semantic annotations and could be resolved via a semantic search, whereas "handles are globally similar in shape" will be resolved by applying a geometric search to the models selected by the semantic search.

To bring intelligence to 3D content and help users creatively manipulate, search, and share 3D objects, we need metaphors for describing the content that are, to some extent, equivalent to those that people employ in organizing their worldview. Humans easily describe and compare objects using shape properties at a conceptual level. However, the digital world requires a more formal method using a variety of computational tools to describe, match, or manage digital objects on the basis of shape properties we can detect, extract, or even quantify, at difference degrees of performance.[5]

Once we solve these problems, we can use 3D to exploit the Semantic Web's capabilities. The most interesting development will occur in scientific and industrial areas, where 3D models are just part of the picture. In computational sciences and applications dealing with 3D data, the relevant scientific data that should be made available to the research community includes not only raw data but also worked-out 3D models; algorithms to implement solutions for specific problems; benchmarks capturing a given task's computational and methodological criticalities; scientific workflows that draw a path between raw measurements, processing, and analysis steps; and the publication of results in traditional scientific journal papers. To develop effective scientific support, we must consider not only networking mechanisms that enable access to scientific facilities or the pooling of scientific data but also approaches to facilitate the sharing of scientific knowledge as a whole to develop a Semantic Web infrastructure that implements access to knowledge via advanced services and middleware layers.[15]

For generic usage scenarios, we must consider that inexperienced users are becoming more actively involved in the content creation pipeline and asking for more intuitive and effective tools for creating, sharing, retrieving, and reusing 3D content. These issues are common to more established media types, such as images and videos, but 3D media open a number of specific challenges due to the geometric nature of the data involved. Even nonprofessionals can easily remix 2D content with software such as Photoshop or FinalCut, but the same level of functionalities for 3D media are available only in complex software systems. ■

Acknowledgments

This work is partially supported by the Focus K3D Coordination Action, EU Contract ICT-2007.4.2, contract number 214993.

References

1. S. Nichols, "Building the 3D Internet," *iTnews*, 2006; www.itnews.com.au/News/NewsStory.aspx?story=42704.
2. M. Greaves, "Semantic Web 2.0," *IEEE Intelligent Systems*, vol. 22, no. 2, 2007, pp. 94–96.
3. J. Hunter, S. Little, and R. Schroeter, "The Application of Semantic Web Technologies to Multimedia Data Fusion within e-Science," *Semantic Multimedia and Ontologies*, Y. Kompatsiaris and P. Hobson, eds., Springer, 2007, pp. 207–226.
4. S. Havemann and D.W. Fellner, "Seven Research Challenges of Generalized 3D Documents," *IEEE Computer Graphics and Applications*, vol. 27, no. 3, 2007, pp. 70–76.
5. M. Spagnuolo and B. Falcidieno, "The Role of Ontologies for 3D Media Applications," *Semantic Multimedia and Ontologies: Theory and Applications*, Y. Kompatsiaris and P. Hobson, eds., Springer, 2007, pp. 185–205.
6. I. Biederman, "Recognition-by-Components: A Theory of Human Image Understanding," *Psychological Rev.*, vol. 94, no. 2, 1987, pp.115–147.
7. D. Hoffman and W. Richards, "Parts of Recognition," *Cognition*, vol. 18, issues 1–3, 1984, pp. 65–96.
8. A. Shamir, "A Survey on Mesh Segmentation Techniques," *Computer Graphics Forum*, vol. 27, no. 6, 2008, pp. 1539–1556.
9. M. Attene et al., "Mesh Segmentation—A Comparative Study," *Proc. IEEE Int'l Conf. Shape Modeling and Applications* (SMI 06), IEEE CS Press, 2006, pp. 14–25.
10. M. Mortara, G. Patané, and M. Spagnuolo, "From Geometric to Semantic Human Body Models," *Computers & Graphics*, vol. 30, no. 2, 2006, pp. 185–196.
11. M. Attene et al., "Characterization of 3D Shape Parts for Semantic Annotation," to be published in *Computer-Aided Design*, 2009; doi: 10.1016/j.cad.2009.01.003.
12. L. De Floriani, A. Hui, L. Papaleo, M. Huang, May and J. Hendler, "A Semantic Web Environment for Digital Shapes Understanding," *Semantic Multimedia*, LNCS 4816, Springer, pp. 226-239, 2007.
13. J.W.H. Tangelder and R.C. Veltkamp, "A Survey of Content Based 3D Shape Retrieval Methods," *Multimedia Tools and Applications*, vol. 39, no. 3, 2008, pp. 441–471.
14. B. Bustos et al., "Content-Based 3D Object Retrieval," *IEEE Computer Graphics and Applications*, vol. 27, no. 4, 2007, pp. 22–27.
15. J. Hunter, "Tracking the Progress of Multimedia Semantics—from MPEG-7 to Web 3.0," *Semantic Multimedia*, LNCS 5392, Springer, p. 1, 2008.

Michela Spagnuolo is senior researcher at the Institute of Applied Mathematics and Information Technologies of the National Research Council of Italy. Contact her at michi@ge.imati.cnr.it.

Bianca Falcidieno is research director at the Institute of Applied Mathematics and Information Technologies of the National Research Council of Italy. Contact her at bianca@ge.imati.cnr.it.


Top articles, podcasts, and more.
cn
computingnow.computer.org