

Towards a new approach for the description of shapes from multimodal data

Z. Li^{*a}, F. Giannini^a, B. Falcidieno^a

^aInstitute for the Applied Mathematics and Information Technologies
CNR

J-P. Pernot^b, P. Véron^b

^bArts et Métiers ParisTech, LSIS - UMR CNRS 7296
Aix-en-Provence, France

Abstract— The construction of a Virtual Environments (VE) requires a long iterative modeling and modification process. Depending on the final purposes, many actors can be involved both in the early conception and in the detailed specification of what has to be included and how it has to be organized. These actors may have different knowledge and expertise. It is therefore important to define tools easy usable even by nonprofessionals in order to facilitate the VE specification and setup. Such tools should allow the incomplete specification of details and the reuse of existing assets and data, either 2D images or 3D models. In this perspective, the work presented in this paper proposes a new shape description model for the management of objects and assemblies, resulting from the combination of multimodal data, together with their intrinsic properties. Based on such a model high level methods and tools directly working on components (either 2D or 3D) can be developed. Thus making easier the solution specification by the application domain experts, which can thus better integrated within the VE design process. This new shape description model is thought to be used as an intermediary model shared between the various actors in the generation process of VEs to keep the link and digital chain between them. The concepts of this model are proposed and illustrated through a first implementation.

Keywords— *geometric modeling, CAD, multimodal data, virtual reality, semantics, shape description, CogInfoCom*

I. INTRODUCTION

Nowadays, VEs are more and more used in different application domains. For example, in architecture area [1], in education domain [2], in product conception [3], etc. In general, the global VE construction process remains the same (Fig. 1). It may involve many actors from different disciplines and expertise in all the definition phase of what has to be included and how it has to be organized. From functional specifications defined by the end user, a certain number of 3D models are generated (or imported from some database) and enriched by various properties (texture, gravity, reflection, etc.) and then they are inserted in the VE with behaviors and functions. This last step requires the programming of each object and of the whole VE. After this step, the end users test the application, and changes are generally required most of the time because of misunderstanding or not sufficiently detailed conceptual specification. Despite the big improvements in the implementation of detailed modelling systems, the modification process of a VE is still very long and tedious. For example, if the form of an object in the scene does not satisfy the end user, changes are applied to the 3D model. As a consequence, the

textures need to be modified, the physical properties have to be recalculated (collision, gravity, light reflection, etc.), the animations may need to be redefined and the related scenarios rescheduled. Under this context, the description of 3D objects in the scene becomes the key problem in the construction of a VE.

Currently, there are no suitable tools to support the conceptual design phase of VEs efficiently. In this phase, the actors should have the possibility to make hypotheses such as: on which assets to consider and how to arrange them possibly taking advantages of already existing resources, etc. They should be able to communicate their ideas easily by mocking up rough and incomplete solutions without requiring deep modelling expertise. It is then crucial to define systems that take into consideration basic cognitive aspects regarding the selection and combination of shapes while guaranteeing model re-editing and processing. In this perspective, it is necessary to develop intuitive user interfaces but also new modelling techniques dealing with multidimensional data (e.g. images and 3D models). The work presented in this paper proposes a new shape description model for the management of objects and assemblies, resulting from the combination of multimodal data, together with their intrinsic properties. Such a model is conceived according to two main purposes. The second aim is to provide high level methods and tools directly working on components (either 2D or 3D) rather than on the low level geometric models. Such a modeling simplification improves the brainstorming activity and lets the application domain expert integrate within the VE design process, which will simplify the

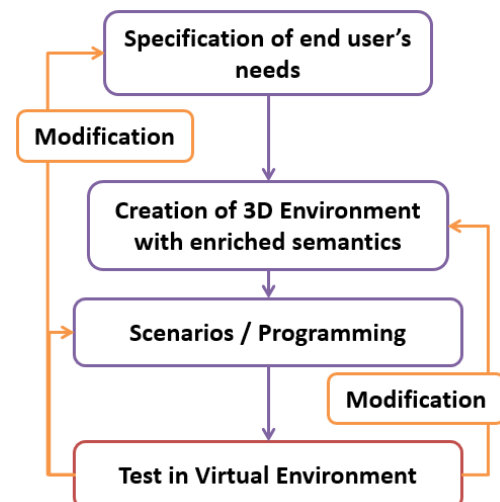


Fig. 1. Global VE construction process.

later on modification process. This new shape description model is conceived to be used as an intermediary model shared between the various actors in the generation process of VEs to keep the link and digital chain within them. Here the underlying concepts of this model are proposed and illustrated through a first implementation.

The paper is organized as follows. Section II is structured in two parts: Section II.A describes the generation process of 3D objects for VEs and motivates the need of a new shape description model from multimodal data; Section II.B lists the necessary characteristics the new shape description model has to satisfy. Then, Section III presents a first idea of this new shape description model according to a multi-layer information approach for shape description. Section 4 shows the initial implementation results. Finally, Section 5 concludes the work presented in this paper and leaves some perspectives for future works.

II. 3D OBJECT GENERATION PROCESS FOR VE

This section describes the current process for generation of VEs together with a possible alternative for a better integration of the application experts in the definition loop. This is obtained through the development of tools allowing an easier and more intuitive idea generation for the wished assets.

A. The digital chain for creation of VEs

Independently on the application contexts considered, i.e. the purposes of the VE created, the generation process is generally following the steps indicated in Fig. 1. The 3D object generation step involves four main actors (see Fig. 2). Each of them has different expertise and perspective on the object. The whole process is mainly a sequential chain with several exchanges among the actors, which have different objectives:

- Expert of the application domain. This is the person demanding a VE to realize some virtual tasks. He/she knows very well the application domain but he/she may not have any knowledge in programming or computer graphics. The tools that he/she can use to express his/her ideas are composed of key words, handmade sketches, pictures or existing 3D models found in certain database. He/She will explain his/her idea of an object or of the whole VE using these multimodal data while indicating the rules of the application domain. Unfortunately, with current tools these multimodal data cannot be used directly into the digital chain.
- Expert of CAD usage: They are product designers who create CAD models from the description of previous actor respecting the technical criteria of the object model purposes. For instance, if the object model represents a product to be manufactured, then such a model will describe all the details of the product needed to its manufacturing (size, surface roughness, tolerance, adjustment, etc.). For realizing virtual tasks in a VE, these models are too heavy. Some information might be not necessary for the VE (the criteria of manufacturing, details of inner parts, etc.) and the resolution of the 3D model might be too high and not adapted to the capacity of the rendering engine for a VE.
- Expert of VE Design: They are the people who simplify the CAD models and associate them with rendering properties (material, texture, light reflection, etc.) to make sure that they

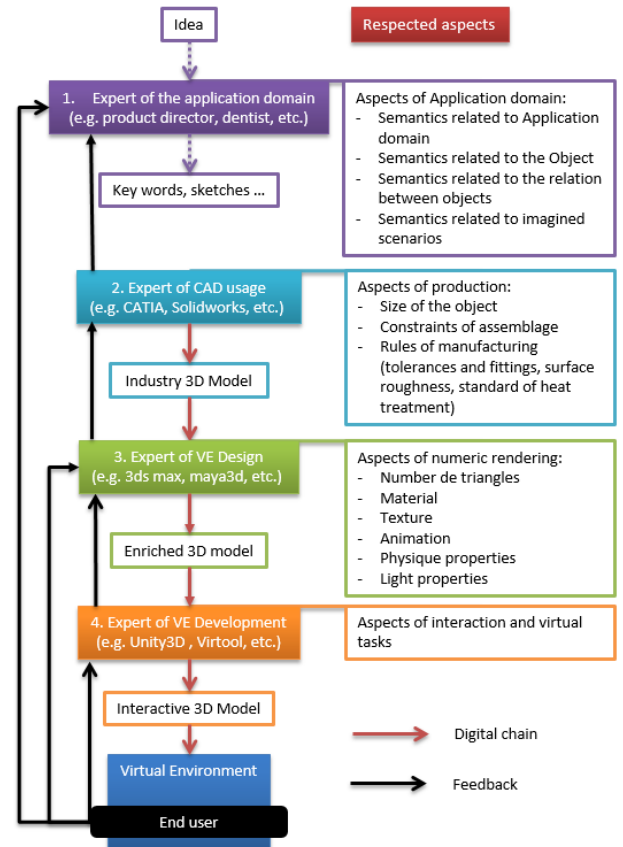


Fig. 2. Classical 3D object generation process.

will be well exposed in a VE. This step is manual, thus breaking the digital chain.

- Expert of VE Development: People in this step add functions and behaviors to each 3D model. First, they need to enrich the models generated by Expert of VE Design with functional properties such as gravity, skin textures, skeleton, animation, etc. Then they add interactions (between objects, between objects and users) and virtual tasks to the object to realize a useful VE finally.

As shown in Fig.2, with the idea given by the expert of the application domain, the digital chain starts from the CAD designer and then pass to the VE designer and the VE developer. In each phase, there is a specific digital model (Industry 3D model, Enriched 3D Model and Interactive 3D model) which follows strictly the aspects (right part in fig.2) defined by their own domains. In the end, the end user will test the VE and provide feedbacks. If he/she identifies that some modifications need to be done on the objects' shape, all the steps, both digital and manual, of the preparation sequence are re-done. The generation process is linear with different feedbacks.

To reduce this modification cycle and facilitate communication between the actors we foreseen a new 3D object generation process for VEs that can limit the break of the digital flow during the process (see Fig. 3).

Compared with the classical process presented in Fig. 2, this process is no longer linear, and feedbacks outside this process are limited. We identify two important key characteristics that permit this process:

- Input of this process is multimodal data (key words, 2D images, sketches, 3D models, etc.). They are easy to be acquired

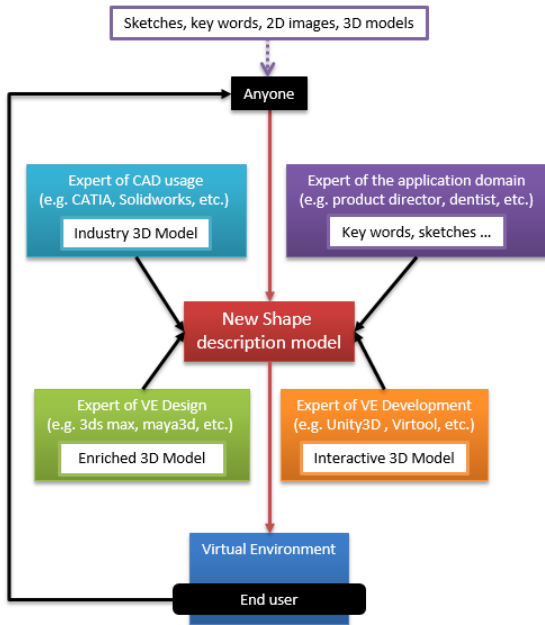


Fig. 3. Proposed new 3D object generation process.

by anyone. Thus, no knowledge is required on specific systems, such as CAD or programming.

- A new shape description model able to support such multimodal data composition, shared by the various actors as a common source of information and data for the process. Thus, this generic model integrates the views of the four actors making easier their collaboration and brainstorming activities than in the classical process.

Those two different 3D object/VE generation processes (classical process & proposed new process) could also be considered as two different Cognitive Infocommunications (CogInfoCom) systems. As the definition shown in [4], we can see the comment points and the differences between them. They are both a complex CogInfoCom system. They both have various levels of cognitive capabilities. In the classical process, the mode of information transferring from the first actor (expert of the application domain) to the second actor (expert of CAD usage) is an inter-cognitive communication. From the second actor to the final VE is an intra-cognitive communication. And between the VE and end user, the mode of information transferring is back to an inter-cognitive communication. In the whole process of the classical system, information is transferred with a representation-bridging communication way. For the proposed process, we have a new shape description model which links the input actor and the final VE together directly using an intra-cognitive communication mode with a representation-sharing communication type. So that the desires of the one who wants to create a 3D object/VE and the need of the one who is going to use the 3D object/VE could be both more easily satisfied. This new shape description model also makes the communication between different actors from the representation-bridging communication type to a representation-sharing communication type as they share the same shape description model.

In conclusion, the new process presented takes multimodal data as input and describes an object with a generic description

model that becomes the reference model for all the various actors working on the VE creation. The process provides more freedom and support the end user expressing their own desires. The specification of the generic shape description model and of the functionalities of its creation and modifications become the central problem to solve.

B. Requirements for the multimodal data representation model

As discussed in Section II.A, a key element to define the new 3D object generation process is the generic shape description model for handling multimodal data. Such a model should present some essential characteristics:

- Independency to the application domain. That means this model can be a reference for any application and actor.
- Capability to manipulate multimodal data. This model should be able to combine, modify and handle all different kinds of multimodal data, namely 2D images, texts and 3D models, while keeping the link between the original data. It should take into consideration the fact that each of them has its own data format, data structure, characteristics and associated semantics.
- Easy to generate by anyone. As discussed in previous section, the one who create this model could be someone who has no any CAD or programming skill. This means that the model should be suitable for developing some high level approaches and tools for the object creation.

To conclude, in this section we showed the classical process for generating a 3D object in a VE and proposed a new process. This new process is based on a generic shape description model, which is generated from multimodal data by anyone and can integrate with the four actors present in the classical process.

III. TOWARDS A GENERIC SHAPE DESCRIPTION MODEL (GSDM)

The first question we need to answer is what the shape of an object is exactly. Which information have to be stored in order to manipulate them easily and meaningfully? In the project called Aim@Shape [5], a new paradigm for understanding and representing shapes has been proposed. Aim@Shape define a shape as any individual object having a visual appearance, which exists in some (two-, three- or higher- dimensional) space. Such a definition includes a great variety of elements, such as pictures, sketches, images, 3D objects, videos, 4D animations, etc. The defined paradigm explains shapes according to three different levels of information: Geometry, Structure and Semantics (GSS). The three levels can be understood as follows:

Geometry level: corresponds to the spatial extent of the object.

Structure level: an abstraction according to the main characteristics of the object, i.e. features and part-whole decomposition

Semantics level: information associated to the object or a part of the object that specifies the meaning or purpose related to the context of use.

This paradigm shows us a new way to understand a shape. It could also set criteria to compare different shape representations/descriptors as shown in the table 1.

In literature, we have different kinds of shape descriptors, some of them are applicable at only one specific level, and others are dealing with two or more. Some of them are focusing on the geometric level of shapes only (mesh, voxels, wireframe, point cloud, etc.), some are strongly linked to special application domain/semantics (Circularity descriptor, rectangularity descriptor, etc.) and some shows the structure of decomposition or partition of the shape (Reeb graph, Skeleton, segmentations, etc.). The existing descriptors and description today are all related to some specified dimension of the data, so they might be not general enough to cover all the three aspects of Geometry, Structure and Semantic for multimodal models.

Therefore, here we propose a new Generic Shape Description Model (GSDM). This model describes shapes according to all the three aspects in the modelling paradigm introduced by Aim@Shape. We also assume that the input multimodal data for the new GSDM approach contain the geometry already enriched with structural information such as part segmentation [6], medial axis [7] and skeletons [8]. We make this hypothesis because many shape analysis approaches [9] already exist today for each type of multimodal data. We consider such a structural decomposition since it allows a meaningful subpart selection for both the composition and positioning thus reducing the cognitive effort from the end user side.

The GSDM can be seen as a hypergraph, whose nodes correspond to *Components* and hyperarcs, i.e. arcs connecting nodes, to *Relations* among them. The scene graph can be built using the same GSDM structure to associate specific information and actions of the VE under construction.

A *Component* is the most basic element in the GSDM and is itself organized according the GSS approach. Different input data will supply different types of “Component”, as shown in table 1.

Multimodal data	Component
text	sentences, key words
2D Images	Segmented part (e.g. pixels group with a label) With enriched information (e.g. media axis, skeleton, etc.)
3D Meshes	Segmented part (e.g. triangle group with a label) With enriched information (e.g. media axis, skeleton, etc.)

Table 1 Examples of Components according to the type of data

Relations are used to link together *Components*.

More than one Relation can link the same Components. Different kinds of Relations are considered specifying the operation to perform. Depending on the type of operation, it can also include Constraints, which indicate the basic elements for the operation. Relation is a complex concept in our GSDM. It crosses all the three levels of the GSS model, see Table 2.

The concept of constraint we are discussing here is not the same as the traditional CAD assembling. The traditional assembling constraints work only on the assembly structure level. The constraints discussed here could work on all the three levels of semantic, structure and geometry and are used for positioning the multimodal elements constituting the final object and guiding the planned operation. We considered four types of

Operation: Shaping, Merging, Union and Grouping. Shaping modifies the geometric, e.g. morphing, bending, subtraction, and possibly the structural layers of information. Merging indicates the traditional union operation in CAD. Thus, it works at all the three GSS layers of the component. Union works on the semantic and the high structure level. It corresponds to put together elements without generating the geometry of the result. Finally Group works only on the semantic level. It is used to logically join components, such as, “the legs of a table”, “the arms of a human body”. These various group relations can be applied depending on different usage contexts.

If we look on the whole scene, then only the Union and Group Relation can be applied to build the structure of the scene, telling us how different objects are positioning the one to the other by adding constraints.

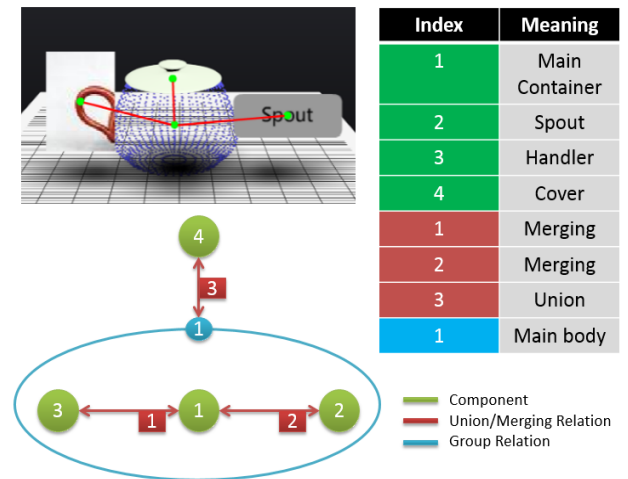


Fig. 4. . Example of GSDM

Acting on information level	Object				Scene	
	Shaping	Merge	Union	Group	Union	Group
Semantic	X	X	X	X	X	X
Structure	X	X	X		X	
Geometry	X	X				

Table 2 The information layer affected by the various types of relations

To make easy the user interaction and relation specification, we take advantage of key entities of the components at structural and/or geometric level, e.g. the symmetry axis or the skeleton elements and points or pixels on them. In general, there are two types of key entity, geometric entity and parametric entity. Geometric entity lays on the local reference frame so it is changed during the transformation of the local reference frame, i.e. it will not follow the transformation or deformation of the Component. Parametric entity is linked to the Component/Components with parameters, it follows the transformation or deformation of the reference frame and the Component/Components depending on the parameters. Key

entity is created automatically by importing a new input data or by user during the design process.

In Fig. 4 an example of GSDM is depicted: green nodes represents Component, red array means the Union/Merging Relation and blue circle shows a Group Relation. The details of this graph are hidden to the user.

IV. FIRST IMPLEMENTATION

To verify the feasibility of our approach, we are implementing a prototype using Unity 3D environment (Fig. 5). In reality, the GSDM model is independent of any system. We selected Unity 3D because it is a powerful engine for the creation of VEs. This implementation works in four steps to realize the generation of the GSDM:

- Import of multimodal data. In this step the user can import any kind of multimodal data into a resources pool (Figure 6-1). Such data should have associated some additional information, i.e. segmentation, medial axis, possibly annotated. A basic GSDM is created for each asset imported.
- Selection of Component. In this step, you can choose one imported multimodal data, and there will open a window shows all the segmented parts that included in this data (Figure 6-2). Then you can choose the parts that you want and add them to the scene.
- Management of key entity. In this step the user can add, modify or delete key entities.
- Build of Relation and Group. In this step the user can build Relations between Components/Groups by applying constraints. User can also group different Components/Groups into a new Group.

There is no a fixed order of these four steps, user can do them separately. And in the end user can display the graph of GSDM in the 3D scene (Figure 6-5). Once the components have been instantiated and the relations specified, the associated optimization problem is solved and the solution is displayed. In the proposed approach, we try to minimize the displacements and deformations between the initial configuration and the one which satisfies the relations.

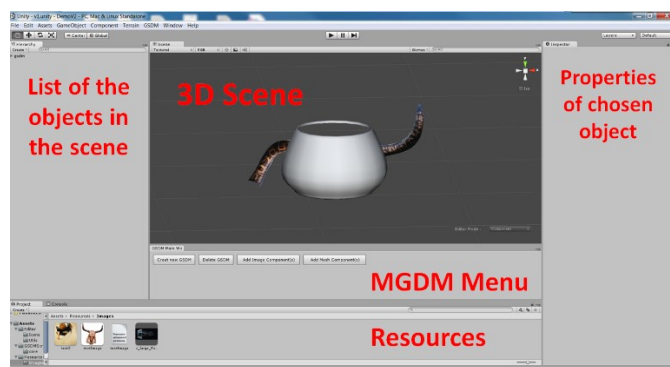


Fig. 5. Overview of the prototype layout

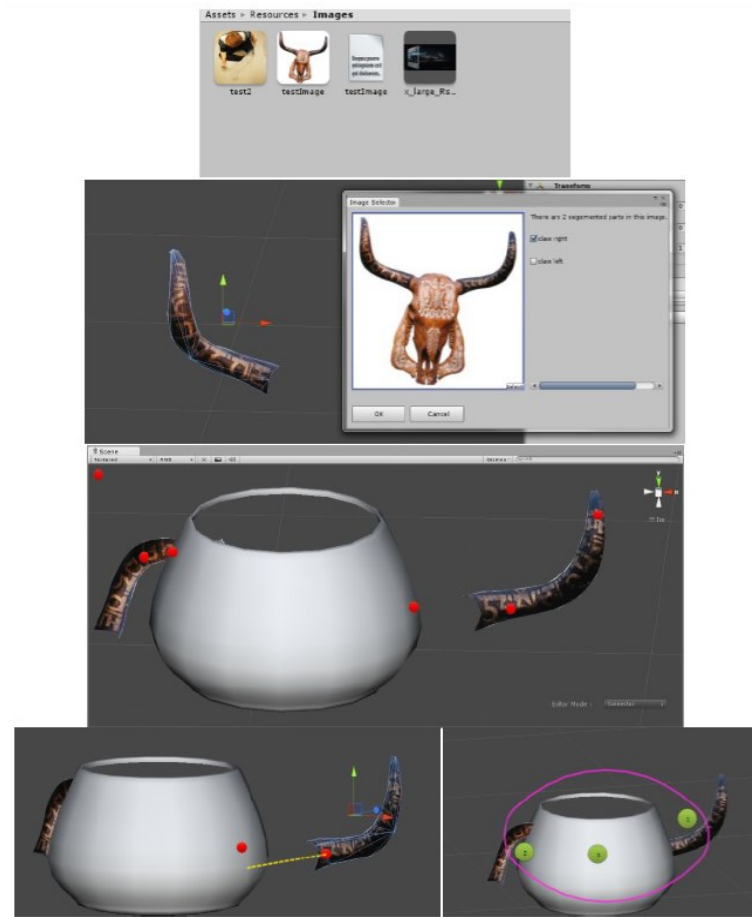


Figure 6: 1. Resources Pool; 2. Choose of Component; 3. Management of key entity; 4. Building constraints; 5. Graph of GSDM

V. CONCLUSIONS AND PERSPECTIVES

In this paper we showed a new 3D object generation process for VEs by proposing the concept of Generic Shape Description Model (GSDM). This model could be used to describe the shape of one object or the whole scene. For now, we are working on the details of the data structure of this model (e.g. different kinds of constraints) and on completing the formalization of the model. For the work in the future, we plan to go on the implementation of the prototype and to solve problems related to the modification of the GSDM: for example, the constraint satisfaction problem, the combination of shape skeleton, etc. Secondly, as the work presented in this paper is a part of the Co-Dive (Conceptual Design in Virtual Environment), we also plan to integrate the modeler with the touch screen user interface under development by ENSAM Cluny partners who are working on the manipulation of objects in the scene and integration mode with VE. As long term future work we plan to integrate, when possible, or to develop algorithms that can transform our GSDM into different representations of the object like a 3D mesh or a rendered 2D picture.

ACKNOWLEDGMENT

The work has been partially supported by the VISIONAIR project funded by the European Commission under grant agreement 262044. It is part of the Co-DIVE project funded by Arts et Métiers ParisTech.

REFERENCES

- [1] A. F. Waly and W. Y. Thabet, "A virtual construction environment for preconstruction planning," *Automation in Construction*, vol. 12, no. 2, pp. 139-154, Mar 2003.
- [2] M. Tassos A. and N. Antonis, "Educational virtual environments: A ten-year review of empirical research (1999–2009)," *Computers & Education*, vol. 56, no. 3, pp. 769-780, April 2011.
- [3] H. Kan, V. G. Duffy and C.-J. Su, "An internet virtual reality collaborative environment for effective product design," *Computers in Industry*, vol. 45, no. 2, pp. 197-213, June 2001.
- [4] Péter Baranyi and Ádám Csapó, "Definition and Synergies of Cognitive Infocommunications" *Acta Polytechnica Hungarica*, vol. 9 no. 1, pp. 67-83, 2012
- [5] "Aim@Shape," [Online]. Available: <http://www.aimatshape.net/>.
- [6] I. Dana E. and W. Paul F. , "Image segmentation based on the integration of colour–texture descriptors—A review," *Pattern Recognition*, vol. 44, no. 10-11, pp. 2479-2501, October–November 2011.
- [7] O. Aichholzer, W. Aigner, F. Aurenhammer, T. Hackl, B. Jüttler and M. Rabl, "Medial axis computation for planar free–form shapes," *Computer-Aided Design*, vol. 41, no. 5, pp. 339-349, May 2009.
- [8] S. Biasotti, D. Attali, J.-D. Boissonnant, H. Edelsbrunner, G. Elber, M. Mortara, G. Sanniti di Baja, M. Spagnuolo, M. Tanase and R. Veltkamp, "Skeletal structures," in *Mathematics and Visualization*, Berlin, Springer-Verlag Berlin Heidelberg, 2008, pp. 145-183.
- [9] D. Zhang and G. Lu, "Review of shape representation and description techniques," *Pattern Recognition*, vol. 37, pp. 1-19, 2001.
- [10] Kszhdan M., Funkhouser T., Rusinkiewicz S. : "Symmetry descriptors and 3d shape matching" in *Eurographics Symposium on Geometry Processing (2004)*, Scopigno R., Zorin D., (Eds.), USA, ACM. 2004, pp. 115-123.
- [11] M. Eugenia Montiel, Alberto S. Aguado, Ed Zalusa, "Surface subdivision for generating superquadrics" *The Visual Computer*, vol 14, no 1, pp. 1-17.
- [12] Kazhdan M., Funkhouser T., Rusinkiewicz S., "Rotation invariant spherical harmonic representation of 3D shape descriptors" in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, Aachen, 2003, pp. 156-164.
- [13] Osada R., Funkhouser, T., Chazelle B., Dobkin D., "Shape distributions" *ACM Trans. Graph.*, vol 21, no. 4, pp. 807-832, Oct. 2002.
- [14] Freek Reingers, Melvin E.D. Jacobson, Frits H. Post, " Skeleton graph generation for feature shape description" in *Data Visualization 2000*, Vienna, Springer Vienna, 2000, pp. 72-83.
- [15] Paul L. Rosin, "Measuring shape: ellipticity, rectangularity, and triangularity", *Machine Vision and Applications*, vol. 14, no 3, pp. 172-184, July 2003
- [16] Martin Reuter, Franz-Erich Wolter and Niklas Peinecke, "Laplace-beltrami spectra as "Shape-DNA" of surfaces and solids", *Computer-Aided Design*, vol. 38, no. 4, pp. 342-366, 2006
- [17] Sitek Arkadiusz, Huesman Ronald H, Gullberg Grant T, "Tomographic reconstruction using an adaptive tetrahedral mesh defined by a point cloud", *IEEE Trans Med Imaging*, vol. 25, no. 9, pp. 1173-9, Sep. 2006.
- [18] Firooz A. Sadjadi and Ernest L. Hall, "Three-dimensional moment invariants", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-2, no. 2, March 1980.
- [19] Tangelder, Johan W. and Veltkamp, Remco C. "A survey of content based 3D shape retrieval methods", *Multimedia Tools Appl.*, vol. 39, no.3, pp 441-471, Sept. 2008.
- [20] Weber G., Bremer P.-T., Day M., Bell J., Pascucci V. " Feature tracking using reeb graphs" in *Topological Methods in Data Analysis and Visualization*, Berlin, Springer Berlin Heideberg, 2011, pp. 241-253.