Optimal Scene Interpretation: Semantic Management of 3-D Objects from a Point Cloud Scene

Paul Cotofrei, Christophe Künzi and Kilian Stoffel

Information Management Institute, University of Neuchâtel, Switzerland
{paul.cotofrei, christophe.kuenzi, kilian.stoffel}@unine.ch

Abstract. This paper presents the main concepts of a project under development concerning the analysis process of a scene containing a large number of objects, represented as unstructured point clouds. To achieve what we called the "optimal scene interpretation" (the shortest scene description satisfying the MDL principle) we follow an approach for managing 3-D objects based on a semantic framework based on ontologies for adding and sharing conceptual knowledge about spatial objects.

1. Introduction

Point clouds are one of the most primitive and fundamental manifold representations. During the lasts years, the accessibility of 3D shapes acquisition devices, such as laser range scanners, facilitated the creation of many 3D geometric databases. These devices typically produce an unstructured cloud of sample points (possibly with noise), where each point encodes information on the shape attributes, such as 3D position, surface normal, surface color, material properties, etc. Applications based on the manipulation and the analysis of such points are extensively used in many disciplines, such as mechanical engineering, architecture, bio-medicine, robotics, but also in other domains such as history and archeology.

The interpretation of a *scene* (query data consisting of partial 3D point clouds of (un)known 3D objects) is normally defined as knowing *which* model is located *where* in the scene. Such an interpretation binds the entities in the scene to the models that we already have the knowledge about. We propose to extend the definition of the "interpretation" task, by including the discovery of spatial relationships between the instances of the models in the scene. This extension allows the acquisition of a new kind of knowledge, concerning the possible repeated, regular patterns of objects spatial distributions. Furthermore, if the set of models and the set of spatial relationships are the elements of a spatial description language, then the concept of *optimal scene interpretation* is well defined, expressing the shortest description (in this language) of the scene in terms of known objects and simplest neighborhood relations between them.

The goal of our project (actually under development) is the establishment of a flexible approach (including framework, methodology, processing methods and finally a working system) allowing the optimal interpretation of a scene (according to our extended definition), containing a large number of objects. To reduce the complexity of the interpretation process in the perspective of the large diversity of real-world situations, the project's framework is based on the following assumptions:

- the scene or the model point clouds are not uniformly sampled nor overlapping;
- the objects of interest are rigid, free-form objects;
- the models exists in the database prior to recognition;
- a description language, based on the models from database and a selected set of spatial relationships, is defined and encoded as a set of fixed ontologies

Stated succinctly, the design of the proposed scene interpretation system (in the following denoted RRR system) involves a three stage processing:

1. *Representation*: The objective is to derive from the point cloud a rich, compact yet meaningful description of the object for efficient storage and for fast and accurate retrieval during recognition.
2. *Recognition*: The derived spatial and geometric descriptions of the partial point cloud from the scene are compared with stored models of objects in order to identify which of those objects are present in the scene. This involves the tasks of instance classification, determination of alignment parameters (rotation, translation) and localization.
3. *Retrieval*: The spatial relationships existing between the objects in the scene are discovered. In a first phase, these relations are analyzed by a pattern finding algorithm to extract possible regular patterns implying the models. In a second phase, a specific ontology describing the scene (which includes as instances the previous discovered relations), is processed to extract an optimal (according to the Minimum Description Length Principle) scene description.

The data we intend to use to check the validity of our approach is a collection of multi-dimensional points having several characteristics along with the spatial location. This data stems from a large project conducted by the Karman Center [1], for which the Pantheon in Rome was scanned (the result is a 3D digital model with more than 620,000,000 points). Therefore, if the performances of the RRR system will be satisfactory, the application will be integrated into the Pantheon project at the disposal of the archaeologists and historians.

The first two processing stage of the RRR system (described in the Section 2) are well documented in the literature (methodologies and algorithms) and therefore, during this project, we intend only to conduct performance -comparison studies in order to select the best solution regarding the data type we analyze (point clouds). On the other hand, the Retrieval phase (detailed in Section 3) represents - in our opinion, after a deep as far as possible bibliography study – an innovative idea which is directly linked to a new approach in computer graphics, the use of techniques and methodologies from Artificial Intelligence and Knowledge Management for scene understanding (see [2], [3], [4]).

2. The Problematics of the Representation and the Recognition Stage

2.1 The Representation Stage

Despite the different application contexts of free-form object models, some criteria apply to representations regardless of domain. According to Brown [5], the general mathematical properties exhibited by object representation schemes are *ambiguity*, *conciseness* and *uniqueness*. The ambiguity or completeness measures the representation's ability to completely define the object in the model space, the conciseness represents how efficiently or compactly the description defines the object, whereas the uniqueness is used to measure if there is more than one way to represent the same object, given the construction methods of representation.

The choice of the object representation is one of the most important decisions for the performance of our RRR system, and must be accompanied by robust techniques for extracting compatible features from both the object model and the input point cloud. Between the two fundamental categories of representation, object-centered and view-centered, the nature of input data and the objective of our project clearly impose techniques from the first category, which attempt to describe the entire 3D volume occupied by the object.

As we already mentioned, the choice of the object model is based on a performance-comparison study of the known various 3D object representations (see [6]) (boundary-based methods [7], volumetric descriptors [8] or spherical representations based on generalized cones [9]) with a particular attention to the capacity to deal with missing data (under-sampling of the surface), with noisy data and with the lacking of connectivity information (unstructured point cloud). Our implicit option is the polygonal mesh representation, especially adapted for point cloud in [10] [11].

2.2 The Recognition Stage

Whereas the problem of finding and identifying objects in single-object scenes with no occlu-

sion has been well studied and many systems designed show good results [12], the same problem, but for multiple objects with the possibility of occlusion and background clutter is much harder. Recognition is performed by matching features derived from the scene with those stored in the model database. Some of the most popular and important approaches to the recognition and localization of 3D objects are *graph matching based on edit distance or graduated asignement*, *interpretive tree search*, *information-theoretic matching*, *hypothesize and test* and *iterative model fitting*.

Given the nature of data, our choice for the recognition process points to the matching algorithm proposed in [13], a thermodynamically inspired algorithm designed to determine a correspondence between the scene and the model point clouds by combining the goodness of the graph-based structural approaches and the entropy-based spatial matching approaches. The maximization of the proposed objective function which captures the structural and spatial differences between point sets, leads to the desired correspondence.

3. The Problematics of the Retrieval Stage

3.1 The Retrieval Stage

A real useful and valuable functionality of an *intelligent* computer vision system would be its capacity to describe an unknown scene as concisely as possible in terms of known objects, transformations of them, and of their mutual spatial relationships. Therefore, we extend the meaning of the scene interpretation process by considering that an *optimal* interpretation of a scene is a description (based on a specific spatial language) which explains the scene in terms of the smallest number of known objects (i.e. known models) and simplest neighborhood relations between them, according to the Minimum Description Length Principle [14]. A simple description language allowing the scene interpretation must include at least rigid, opaque 3D objects, and a set of spatial relationships. From a technical viewpoint, our approach is to "encode" such description language inside a dynamically created semantic layer, added to our 3D point cloud, and expressed as a set of ontologies (in the following denoted as the *reference ontology*) comprising the description of different systems and representation models that might be used.

A reasoning engine processes the low-level knowledge structures captured in the reference ontology. The goal of this reasoning is to deduce new, high-level knowledge and to signal inconsistencies in the conceptualizations. Two main approaches can be applied: using general logic based inference engines or using specialized algorithms (Problem Solving Methods). The logic based inference engines may be classified[1] by the expressivity of the logic they can reason with, from Higher Order Logics (HOL) down to different subsets of First Order Logic (including fuzzy or probabilistic approaches). In general, more expressive logics are more difficult to reason with, where in the worst case scenario there exist no strategies that could ensure the termination of the reasoning process. Concerning the second approach, each PSM represents a declarative, reusable description of reasoning for solving a particular type of problem.

Based on the information learned during the recognition stage (objects instances found in the scene and their exact localization), the system send queries to the reasoning engine concerning the possible binary spatial relationships between the found instances. The set of queries evaluated as true (positive queries) together with the set of object instances in the scene form then the raw data from which the optimal scene interpretation is generated. According to the MDL Principle, the optimal description minimizes the length of the set {*theory*, *data encoded using the theory*}. Consequently, to different types of theories corresponds different optimal descriptions, even if all are based on the same data. In our opinion, two types of theories must be considered as appropriate:

- *Rules*. Significant rules between the classes (models) of the object's instances are extracted using the spatial association rules approach [15]. Another approach, derived from similar applications on natural language [16], generates the grammatical rules

[1] http://www.semanticweb.org/inference.html

underlying the language described by the set of positive queries (considered as correct sentences) using a semi-supervised learning algorithm. The optimal description minimizes the length of rules, together with the exceptions to these rules.

- *Ontology*. As one of the form expressing the paradigm of concept formation (together with clustering or Concept Lattices), an ontology is a conceptual structure used for concisely characterizing data. If we define the length of an ontology as the number of nodes plus the number of links of type *is-a* or *has-a*, then an algorithm for creating an ontology, starting from the set of positive queries (statements) and using MDL as guiding principle, can be designed [17]. The algorithm must iteratively search analogies or isomorphism in *contiguous* sets of statements, where two statements are connected if they share a common symbol (instance or spatial relation), and a set of statements is contiguous if there's a path within the statement set from every statement in it to every other statement in the set.

Each type of optimal description (based on rules or on ontology) has its own advantages/drawbacks regarding the comprehensibility of the final result (an ontology is more adapted for a visual representation than a set of rules), and therefore must be chosen according to the interest of the final user of the RRR system.

As we already mentioned, the description language is encoded as a semantic layer over the raw data recorded as unstructured clouds of sample points. For architectural data (the Pantheon project) we defined [18] a system based on two components - an efficient storage module for 3D data and a concept-based representation module. The second module (detailed in the next subsection) is in fact the reference ontology designed to support the semantic integration of architectural structures (often based on very complicated models, taking into consideration technical and aesthetic aspects).

3.2 The Reference Ontology

The semantic layer is mainly composed of two groups of ontologies: the upper (basic) group and the lower (user) group. This difference between the used ontologies has been shown already in [19]. Without losing in generalization, the user can describe a spatial object in a specific environment by actually constructing the 3D object from elementary shapes. Each elementary shape is described mainly by a transformation (scaling, translation, rotation), one or more positions and one or more dimensions. Since each transformation can be expressed in different ways or is shape–dependent, the upper ontologies comprise the description of different systems and mathematical models that might be used [18]:

1. The *Coordinate Systems Ontology*. This ontology defines a few systems for describing a position in space: cartesian, spherical and cylindrical — the ontology being easily extensible with other systems. Each coordinate system has properties that map its specific characteristics: e.g. the CartesianSystem has three length–based properties (corresponding to the x– , y– and z– coordinates), while the CylindricalSystem has two metric–like properties and a degree–like property that correspond to the radial, vertical and azimuth values, respectively.
2. The *Transformation Systems Ontology*. The same approach has been used for the transformations ontology, i.e. each instantiable rotation system has predefined attributes (e.g. roll angle, vector, etc.) that match their

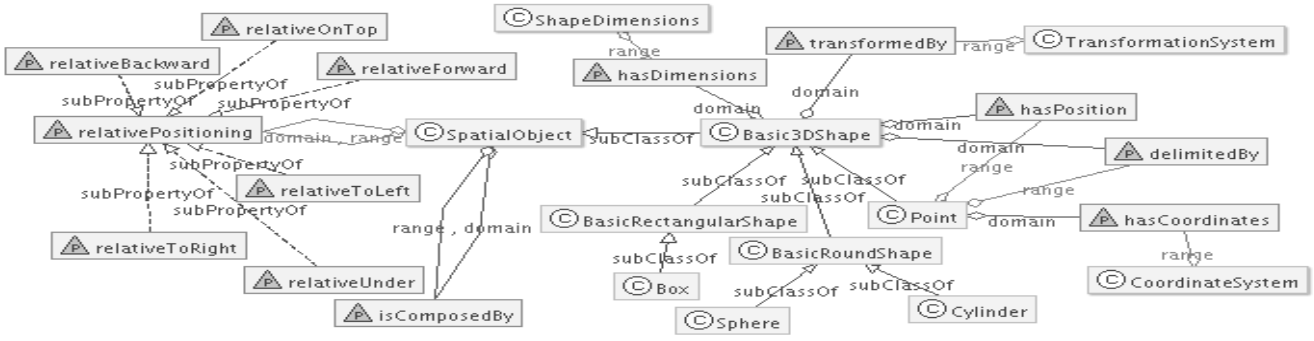

Figure 1. Excerpt from the Geometrical Shapes Ontology

corresponding mathematical elements. For example, the EulerAxisRotation defines properties for rotation vector and angle, while the TaitBryanRotation has a degree–like property for each dimension.

3. The *Geometrical Shapes Ontology*. Inspired by [20], the shapes ontology is the most complex one and it formalizes the fundamental geometrical shapes such as cuboids, sphere, etc. The central concept of this ontology is the SpatialObjet, all basic shapes as well as any user–defined spatial object being subclasses or instances of the SpatialObjet concept. In the spatial ontology, each shape is described mainly by a transformation (e.g. rotatedBy), a position (hasPosition) and by its dimensions (hasDimensions), or it can be identified by one or more points of reference (definedBy). As can be seen in Fig. 1, the spatial relationships between objects are expressed as a hierarchy of properties (RelativePositioning), but it's also possible to be defined as a distinct ontology.

For all of the upper ontologies, the system considers that two parameters are implicit: the distance unit expressed in meters and the degree unit in radians.

The topological and compositional properties defined on SpatialObjet's let the user to construct iteratively more complex SpatialObjet's. When he starts working with the initial system the user can essentially use Basic3dShape and its associated basic operations to define his queries that correspond to its basic objects. By composing these simple Basic3dShape objects, the user can describe new, more complicated shapes. These new spatial objects have to be defined in the *User Ontology*. We will illustrate now an example of how a user might proceed to define its own objects and make them available as an ontological definition. Let's imagine for this example that we would like to find Corinthian columns in the Pantheon data.

By looking at the image of the entrance (see Fig. 2), one user may try to retrieve the points defining a column using the basic definition of a *box*, another user may prefer to retrieve the same points using a *cylinder*, whereas a third user may realize that a combination of the two previous approaches might be more appropriate.

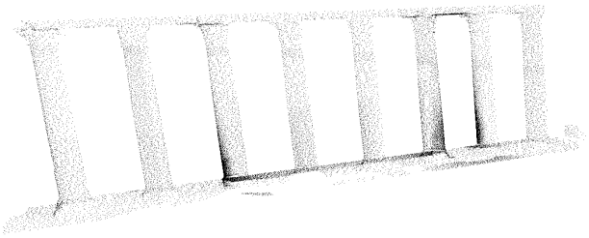

Figure 2. The entrance of the Pantheon in Rome

Let's say he would use a box for the base element, then a cylinder for the middle part of the column, and another box for the top of the column, all of them being combined to define a Corinthian column. For these types of complex shapes, a new concept can be added to the user ontology, named *CorinthianColumn*. Furthermore, another concept *CorinthianEntrance* can be defined as composed of *CorinthianColumn*'s. The *Basic3dShape*'s used to define the new concepts have precise coordinates and ontological descriptions.

Based on the extended ontology another user could add his own concepts and make them dependent on the newly introduced concepts of the *CorinthianColumn*. As known from the history of architecture, Corinthian columns might consist of identical base and middle element, but they could differ in their top element. The ontological definitions should also allow this refinement of the basic definitions of a *CorinthianColumn*.

Once the user ontology is completely populated, the Retrieval phase may be applied in order to obtain the optimal scene interpretation. The system's functionalities for this phase are not yet implemented, but we may illustrate what it could be considered as an optimal description: for the scene representing the entrance, this could be the set including the rule "$instance_1$ of CorinthianColumn *relativeToLeft atDistance d* from $instance_2$ of CorinthianColumn", together with the instance of the first column (satisfying the rule) and of the last column (the exception for this rule).

4. Conclusions

Following a novel direction in computer vision during the last years - the use of techniques

from Artificial Intelligence and Knowledge Management for scene understanding – we started the development of a complex project designed to generate an optimal scene interpretation starting from 3D unstructured point clouds. The novelty of our approach is given, in our opinion, by the definition of the concept *optimal interpretation*, seen as the shortest description (according to the MDL principle) of a scene, expressed in a spatial language encoded as a dynamically created semantic layer, implying 3D object instances and spatial relationships. For architectural data (the Pantheon Project) we designed this layer as a semantic framework for adding and sharing conceptual knowledge about spatial objects, by starting from a reference ontology that describes the basics of the spatial aspects.

Bibliography

1. **The Institute of Advanced Study in the Humanities and the Social Science.** http://www.karmancenter.unibe.ch/karman-center/the_projects/pantheon. [Online]

2. **Plemenos, D.** Using Artificial Intelligence Techniques in Computer Graphics. *International Conference Graphicon.* 1999.

3. **Golfinopoulos, V., Miaoulis, G. and Plemenos, D.** A semantic approach for understanding and manipulating scenes. *Int. Conf. 3IA, Limoges (France).* 2005.

4. **Amitha Perera, A. G., et al.** Moving Object Segmentation using Scene Understanding. *Conf. on Comp. Vision and Patt. Recogn.* 2006.

5. **Brown, C. M.** Some mathematical and representational aspects of solid modeling. *IEEE Trans Pattern Anal. Mach. Intell.* 1981, Vol. 3, pp. 444 - 453.

6. **Campbell, J. R. and Flynn, P. J.** A survey of free-form object representation and recognition techniques. *Comp. Vision and Image Underst.* 2001, Vol. 81, pp. 166-210.

7. **Besl, P.J.** Surfaces in Range Image Understanding. Springer-Verlag, 1988.

8. **Suk, M. and Bhandarkar, S.M.** Three Dimensional Object Recognition from Range Images. Springer-Verlag, 1992.

9. **Ikeuchi, K. and Hebert, M.** Spherical Representations: From EGI to SAI. *Object Representation in Computer Vision.* 1992, pp. 327-345.

10. **Oblonsek, C. and Guid, N.** A fast surface-based procedure from object reconstruction from 3d. *Comput. Vision Image Understand.* 1998, Vol. 69, pp. 185-195.

11. **Jagannathan, A. and Miller, E.L.** 3D Surface Mesh Segmentation Using A Curvedness-Based Region Growing Approach. *IEEE Trans. on Patt. Anal. and Machine Int.* 2007, Vol. 29, 12, pp. 2195-2204.

12. **Murase, H. and Nayar, S. K.** Visual learning and recognition of 3-d objects from appearance. *Int. J. Comput. Vision.* 1995, Vol. 14, pp. 5-24.

13. **Jagannathan, A. and Miller, E.L.** Unstructured 3D Point Cloud Matching within Graph-theoretic and Thermodynamic Frameworks. *Proc. of CVPR.* 2005.

14. **Rissanen, J.** A universal prior for integer and estimation by minimum description length. *Ann. of Stat.* 1982, Vol. 11, pp. 416– 431.

15. **Koperski, K. and Han, J.** Discovery of spatial association rules in geographic information databases. *Proc. of 4th Int. Symp. on Large Spatial DB.* 1995, pp. 47-66.

16. **Petasis, G., et al.** Learning context-free grammars to extract relations from text. *Proceedings of ECAI.* 2008, pp. 303-307.

17. **Pickett, M., and Oates, T.** The Cruncher: Automatic concept formation using minimum description length. *Proceedings of the 6th Int. Symp. SARA.* Springer-Verlang, 2005.

18. **Ciorascu, C., Künzi, C. and Stoffel, K.** Overview: Semantic Management of 3d Objects. *SAMT Conference, S-3D Workshop Proceedings.* 2008.

19. **Grenon, P. and Smith, B.** SNAP and SPAN: Towards Dynamic Spatial Ontology. *Spatial Cognition and Computation.* 2004, Vol. 4, 1, pp. 69–104.

20. **Perry, M., Hakimpour, F. and Sheth, A.** Analyzing Theme, Space and Time: An Ontology-based Approach. *Proceedings of 14th Int. Symp. AGIS.* 2006, pp. 147-154.