Computers & Graphics 33 (2009) 280–290

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Semantics-driven best view of 3D shapes

Michela Mortara *, Michela Spagnuolo

CNR IMATI Ge, via de Marini 6, 16149 Genova, Italy

ARTICLE INFO

Article history:
Received 13 December 2008
Received in revised form
2 March 2009
Accepted 3 March 2009

Keywords:
Viewpoint selection
Semantics
Segmentation
3D shapes

ABSTRACT

The problem of automatically selecting the pose of a 3D object that corresponds to the most informative and intuitive view of the shape is known as the *best view* problem. In this paper we address the selection of the best view driven by the meaningful features of the shape, in order to maximize the visibility of salient components from the context or from the application point of view. Meaningful features can be automatically detected by means of semantic-oriented segmentations: we tested several approaches with very pleasant results in the automatic generation of thumbnails for large 3D model databases.

1. Introduction

The problem of automatically selecting the pose of a 3D object that corresponds to the most informative and intuitive view of the shape is known as the *best view* problem. In many applications, like the creation of thumbnails for huge repositories or catalogues of 3D models, it is necessary to capture a pleasant and informative image of an object; also, informative 2D views of 3D objects can be used to apply various computer vision techniques in the 3D setting, for instance for shape recognition and classification [18]. Up to now, such snapshots are still manually captured with a quite time-consuming process.

Judging a set of views as the best ones for an object is obviously deeply related to the nature of human perception of 3D shapes, which has been matter of study for centuries and is still under debate in various disciplines like neuroscience, psychology and computer science. There is an agreement, however, that the factors that concur to the selection of a canonical view include: *salience and significance* of features for the observer; *stability* of the aspect with respect to small transformations; *aesthetic* criteria, such as preservation of geometric proportions; *familiarity*, that is, for a known object, the view which corresponds to our normal view of the object in the real world is preferred; *functionality*, views that focus on functional parts and suggest how to use or interact with an object are preferred.

In other words, there is an agreement that the best view is closely related to the semantics of a shape and/or of its salient components, in a specific context or application domain, and it should make it easier to recognize the 3D object according to its meaning or purpose. Nonetheless, the majority of prior work on this topic deals with just geometric properties of the object surfaces, and makes use of semantic information only implicitly.

More precisely, methods that addressed the best view selection approach the problem with the definition of a scoring function that is evaluated on a set of admissible viewpoints: the viewpoint which gets the highest score is the one that defines the best view. The scoring function may take into account different characteristics that vary from the bare percentage of visible points from the viewpoint to more sophisticated functions. To the best of our knowledge, however, except for [22] which takes into account global shape characteristics (volume isosurfaces and their topological properties), the proposed scoring functions only capture local, geometric characteristics of the shape and mostly disregard structural aspects, and even more, semantics in the sense of the presence of meaningful components of an object in a given context. Fig. 1 compares the best view of David's head computed using two geometric approaches (the surface curvature and the mesh saliency) to our semantics-driven approach.

In this paper, we argue that the best view should be evaluated taking into consideration the meaningful components, or features, of a 3D shape, that is, the quality of a view should be related to the semantics of the displayed features. In particular, we propose a new scoring function for selecting the optimal view point which blends visibility criteria with a reasoning on the proportion of relevant features which are visible in a given direction. A preliminary version of the results described in this paper was presented in [16].

For our experiments, we tested a number of methods which target the recognition of different sets of features and are suited to different type of shapes. As pointed out in [20], indeed, there is

* Corresponding author. Tel.: +39 0106475677; fax: +39 0106475660.
E-mail address: michela.mortara@ge.imati.cnr.it (M. Mortara).

0097-8493/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cag.2009.03.003

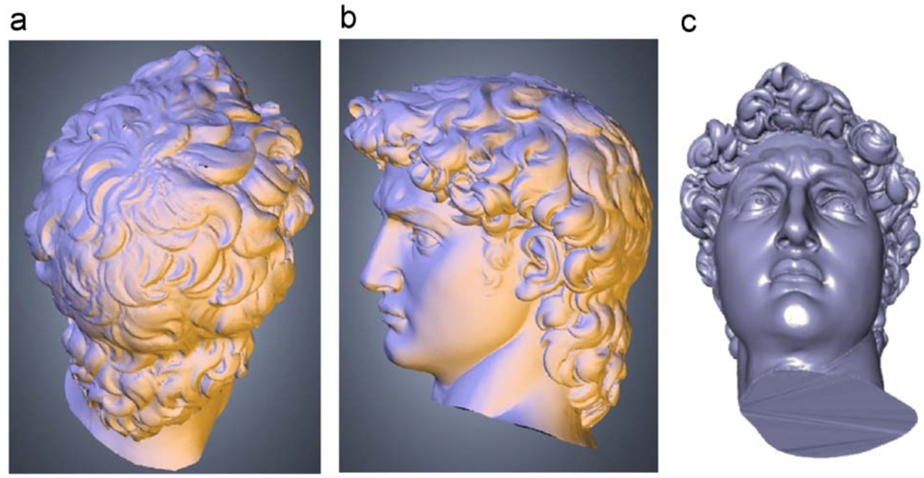

Fig. 1. Best view of David's head computed using the surface curvature (a), the mesh saliency (b) and our semantics-driven approach (c).

not a unique segmentation approach that always gives the best solution: in turn, many algorithms exist in the literature [3] with characteristics that make them particularly suitable on some kind of shapes and inappropriate for others. Therefore, it is a shared opinion in the computer graphics community that only by combining different approaches we will be able to capture the relevant features of objects depending on the user needs, in either an automatic or a computer-assisted fashion. Recently, shape annotation tools have been proposed that exploit a mixture of segmentation algorithms to support the user in selecting meaningful parts [10]. Moreover, tools for the automatic propagation of the segmentation into salient parts from a representative shape to all the similar shapes in a class are likely to appear in the near future.

The core of the proposed approach is the development of a new feature-based viewpoint scoring function that privileges the visibility of meaningful parts which, along with the choice of the most suitable segmentation for a given class of 3D shapes, makes it possible to automatically capture informative images of shapes from the functional and/or aesthetic and/or familiarity point of view. We tested our algorithm over a database of 400 3D models getting excellent results (more than 97% of the views were perceived as good or at least pleasant).

The remainder of this paper is organized as follows: related works on the problem of the best view selection are briefly reviewed in Section 2. Some considerations about the problem of selecting salient features and the brief description of the segmentation algorithms chosen for experimentation are given in Section 3. The core of our approach is detailed in Section 4, and the evaluation of the achieved results is described in Section 5. Discussion about time complexity, view stability and comparison with related works is given in Section 6; finally some plans about future directions conclude the paper.

2. Related work

The choice of a single view to depict a 3D object that can be judged as *the best one* is related to the nature of human perception, matter of study for centuries and still under debate in various disciplines like neuroscience, psychology and computer science. A solid work on this topic is [24], which investigates preferred or canonical views for 3D objects; in particular, the authors point out the factors that concur to determine a canonical view, that are: salience and significance of features for the observer; stability of the aspect with respect to small transformations; number of occluded features; aesthetic criteria (geometric proportions); familiarity, i.e. for a known object the view that is encountered more frequently is preferred; functionality, i.e. views that are more related to how to use or interact with an object are preferred.

While the best view is strictly related to the semantics of the shape and/or to its salient components, the majority of prior works in this topic deals just with geometric properties of the object surfaces, maybe due to the previous immaturity of theories and tools to extract, associate and manage semantics related to 3D digital content.

In most of the approaches, a set of admissible viewpoints is first defined; usually the admissible viewpoints are sampled over a sphere surrounding the object, the *viewing sphere*, see Fig. 3; the choice of a pre-defined set of candidates is needed to reduce the search space of the best viewpoint. An approach different from the uniform sampling of the viewing sphere is proposed in [17] and is based on normal clustering: each normal vector from a vertex on the object surface, in fact, determines a point on the viewing sphere; by clustering these points and taking the centre of mass as a representative of each cluster a set of candidate viewpoints is defined.

Given the admissible viewpoints, the best view problem is solved by assigning scores to them which accumulate information about local shape characteristics that vary from the bare percentage of visible points from that viewpoint to more sophisticated functions that measure the exposed amount of geometric complexity of a shape, basing on properties like surface area entropy, visibility ratio, curvature entropy and others (see [17] and references therein).

A recent and quite popular work in this category is [8], where a saliency measure is introduced to detect perceptually relevant portions of the shape. The saliency measure builds on the computation of the mean curvature at each vertex of the mesh, which is weighted with a Gaussian filter at two different scales. The scales correspond to vertex neighbourhoods of different size, and the saliency is defined as the absolute difference between the Gaussian-weighted averages of the mean curvature at fine and coarse scales. Then, the best viewpoint is the one which maximizes the sum of the saliency for visible regions of the object. In order to avoid an exhaustive search, the authors use a gradient-descent-based optimization heuristic to select good viewpoints on the viewing sphere. This saliency measure captures regions that are different from their surrounding context, and, thanks to this property, it gives good results and does not emphasizes regions withrepeated detailed patterns of features.