Perceptual Information-Theoretic Measures for Viewpoint Selection and Object Recognition

Author: Xavier BONAVENTURA BRUGUÉS

2015

Doctoral Programme in Technology

Advisors: Dr. Miquel FEIXAS FEIXAS
Dr. Mateu SBERT CASASAYAS

This manuscript has been presented to opt for the doctoral degree from the University of Girona
List of publications

Publications that support the contents of this thesis:


Additional publications indirectly related with this thesis:


List of Figures

1.1 Lady of Elche ................................................. 2
2.1 Plot of binary entropy ........................................... 8
2.2 Venn diagram of a discrete channel ............................... 9
2.3 Visibility channel ................................................. 14
3.1 The best view and the corresponding sphere of viewpoints of four different models using $I_1$, $I_2$, and $I_3$ ................................. 20
3.2 The worst view and the corresponding sphere of viewpoints of four different models using $I_1$, $I_2$, and $I_3$ ................................. 21
4.1 Set of best views for the armadillo, the cow, and the dragon selected by the 26 human subjects of the Dutagaci et al. benchmark ................................. 38
4.2 Plot of the error for each method running the Dutagaci et al. benchmark ................................. 39
4.3 The best view and the corresponding sphere of viewpoints of three models using different methods: # visible triangles, projected area, Plemenos & Benayada, visibility ratio, viewpoint entropy / $I_2$, viewpoint Kullback-Leibler, viewpoint mutual information ($I_1$), $I_3$, silhouette length, silhouette entropy, silhouette curvature, silhouette curvature extrema, Stoev & Straßer, maximum depth, depth distribution, instability, depth-based visual stability, curvature entropy, visual saliency, projected saliency, and saliency-based EVMI ................................. 40
5.1 Shape of function $\rho(d)$ ........................................... 46
5.2 Obscurances in videogames ........................................ 47
5.3 Visualization of obscurances, polygonal information $I_1$, polygonal information $I_2$, and polygonal information $I_3$ for the lady of Elche and the angel models ........................................... 51
5.4 Visualization of polygonal information $I_2$ for the lady of Elche and polygonal information $I_3$ for the coffee cup with and without polygonal interpolation ........................................... 51
5.5 Grey-map representation of Tsallis polygonal information $I_1$ with $\alpha=0.6$, and Tsallis polygonal information $I_2$ with $\alpha=0.6$ ........................................... 53
5.6 Grey-map representation of Tsallis polygonal information $I_2$ with $\alpha=0.6$ ........................................... 52
5.7 Visualization of obscurances, Tsallis polygonal information $I_1$ with $\alpha=0.6$, and Tsallis polygonal information $I_2$ with $\alpha=0.6$ ........................................... 53
5.8 Combination of a textured model with Tsallis polygonal information $I_2$ with $\alpha=0.6$ ........................................... 54
5.9 Best views for the viewpoint information $I_1$, viewpoint information $I_2$, $VQ_1$, $VQ_2$, and $VQ_3$ ........................................... 55
### List of Figures

5.10 Worst views for the viewpoint information $I_1$, viewpoint information $I_2$, $VQ_1$, $VQ_2$, and $VQ_3$ ............................................ 56

5.11 Six best views of the lady of Elche using the $I_1$-based algorithm and the $EI_1$-based algorithm weighted by the polygonal information $I_2$ and $I_3$ .... 58

5.12 Six best views of the angel using the $I_1$-based algorithm and the $EI_1$-based algorithm weighted by the polygonal information $I_2$ and $I_3$ .... 59

5.13 Exploratory tour with the extended viewpoint information $EI_1$ weighted by the polygonal information $I_2$ .... 60

5.14 Exploratory tour with the extended viewpoint information $EI_1$ weighted by the polygonal information $I_3$ .... 60

6.1 Two different views of the Book Cliffs terrain model ......................... 65

6.2 Grey-map representation of polygonal information $I_1$, $I_2$, and $I_3$ raised to 1, 2, and 4 ......................................................... 67

6.3 Combination of a textured terrain with polygonal information $I_1$, $I_2$, and $I_3$ powered to 4 ......................................................... 68

6.4 Image enhancement of the original image in 3D space using polygonal measures and in image space using GIMP .................................. 69

6.5 Two different views of the same terrain model enhanced with GIMP with equalize option where some parts are painted with slightly different colors 69

6.6 Six best views of the Book Cliffs terrain from the sphere of viewpoints using $VQ_1$, $VQ_2$, and $VQ_3$ and discarding triangles seen with a quality better than 50% ......................................................... 71

6.7 Six best views of the Book Cliffs terrain from the set of views provided by the helicopter using $VQ_1$, $VQ_2$, and $VQ_3$ and discarding triangles seen with a quality better than 50% ......................................................... 71

6.8 Information gained with each new viewpoint selected for $VQ_1$, $VQ_2$, and $VQ_3$ using the sphere of viewpoints and the set of views provided by the helicopter to texture the 3D model ......................................................... 72

7.1 $I_1$-spheres and $I_2$-spheres corresponding to four different 3D models of the same class ................................................................. 79

7.2 3D models of the same class with similar value of $I(V; Z)$ .................. 81

7.3 Two objects with similar $I(V; Z)$ values and different patterns for the $I_1$-spheres ................................................................. 84

7.4 3D models of the same class where we can see some models where the value of $I(V; Z)$ is quite different between them .................. 84

7.5 Malformed model where we can see the back face of some polygons in black and the background through some holes in white ................. 85

7.6 $I_1$- and $I_2$-histogram corresponding to two models of the same class .... 85

7.7 $I_1$- and $I_2$-histogram corresponding to two models of different classes .... 86

7.8 3D models used for the histogram figures ..................................... 86

7.9 The satellite dish class of the test set ......................................... 86
## List of Tables

3.1 Number of polygons of the models used and computational cost of the preprocessing step for each model in milliseconds ........................................... 21

4.1 The most relevant notation symbols used in the chapter .................................. 27

4.2 List of measures with the corresponding names used in surveys of Polonsky et al., Dutagaci et al., and Secord et al., whether the best viewpoint corresponds to the highest or the lowest measure value, whether the measure is sensitive to the polygonal discretization or not, and the main reference of the measure ......................................................... 28

4.3 Field of application of different viewpoint quality measures and the corresponding reference of the paper ................................................................. 41

5.1 Number of polygons of the models used and computational cost of the preprocessing step for each model in milliseconds ........................................... 50

7.1 For each measure, the size of the signature, the time to generate it, and the time to compare two models are shown ....................................................... 82

7.2 Results of our measures with the Princeton Shape Benchmark training set .......... 83

7.3 Results of our measures with the Princeton Shape Benchmark test set .......... 83

7.4 Results from Shilane et al., where we have merged the results of our $I_1$-sphere method ................................................................. 87
La finalització d’aquesta tesi doctoral tanca una etapa important de la meva vida on m’agradaria tenir present tota la gent que m’ha ajudat a fer-la possible. Primer de tot vull agrair als meus directors, en Mateu Sbert i en Miquel Feixas, per tota la feina, idees i ajuda que m’han donat dins i fora del doctorat i durant el camí previ a aquest. També vull agrair a tots els companys de despatx, de grup, d’edifici o d’universitat que en algun moment han ajudat a fer més plàcid el camí: Ester, Marc, Roger, Màrius, Milán, Olga, June, Rubén, Nico, Anton, Yago, Pau, ... I would like to thank all the people who made possible and pleasant my research stays in the Max Planck Institute for Biological Cybernetics: Lewis, Christian, Farid, Nina, Frank, Stefan, Björn, Florian, Jo, Heinrich, Isabelle, Joost, Janina, Burak, ... I would also like to thank Weiliang, Xiaopeng, and Jianwei because without them it wouldn’t have been possible my research stay in the Chinese Academy of Science in Beijing. I don’t want to forget to thank Ola, Simon, and John for their contributions to my work.

Agraeixo la paciència, suport i comprensió de la meva estimada Laura que només d’arribar a casa ja sabia si els resultats de la tesi eren satisfactoris o no. A en Pou, aquest company de viatge que fa un parell d’any es va unir a la família i que ha fet les seves aportacions a aquest treball tot passejant-se per sobre del teclat de l’ordinador. Als meus pares per tota l’educació que m’han donat al llarg de la meva vida i el seu suport incondicional en les bones i males decisions. També a la meva germana Laura i a tota la meva família, avis, tius, cosins, sogres i cunyats que sempre han estat molt comprensius amb la meva feina. Vull agrair també a tots els meus amics, especialment en Joan, l’Adri i en Crous, que durant molts anys han hagut de patir les conseqüències d’aquesta tesi i m’han ajudat a desconnectar. No em voldria deixar d’agrar als companys de l’AVAP, la Taverna de Jabba, els Aligots, Via Lúdica i tots els que amb el vostre gra de sorra heu fet possible aquesta tesi moltes vegades sense saber-ho.
Acknowledgements

The work in this thesis has been supported by the grant number BES-2011-045252, TIN2010-21089-C03-01, TIN2013-47276-C6-1-R, EEBB-I-13-06602, and EEBB-I-2012-04799 of the Ministry of Economy and Competitiveness (Spanish Government), by grant number 2009-SGR-643 and 2014-SGR-1232 of Generalitat de Catalunya (Catalan Government), and by the National Natural Science Foundation of China (Nos. 61331018, 61372190).

The Book Cliffs terrain model has been provided by Simon John Buckley, Uni Research CIPR, Bergen, Norway.
## 5 Information Measures for Object Understanding

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>Introduction</td>
<td>43</td>
</tr>
<tr>
<td>5.2</td>
<td>Background</td>
<td>44</td>
</tr>
<tr>
<td>5.2.1</td>
<td>Tsallis Information Measures</td>
<td>44</td>
</tr>
<tr>
<td>5.2.2</td>
<td>Obscurances and Ambient Occlusion</td>
<td>45</td>
</tr>
<tr>
<td>5.3</td>
<td>View-Based Polygonal Information</td>
<td>47</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Shannon Polygonal Information</td>
<td>48</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Tsallis Polygonal Information</td>
<td>49</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Results</td>
<td>50</td>
</tr>
<tr>
<td>5.4</td>
<td>Viewpoint Selection and Object Exploration</td>
<td>54</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Viewpoint Selection</td>
<td>54</td>
</tr>
<tr>
<td>5.4.2</td>
<td>N Best Views and Object Exploration</td>
<td>56</td>
</tr>
<tr>
<td>5.5</td>
<td>Conclusions</td>
<td>61</td>
</tr>
</tbody>
</table>

## 6 Information Measures for Terrain Visualization

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1</td>
<td>Introduction</td>
<td>63</td>
</tr>
<tr>
<td>6.2</td>
<td>Background</td>
<td>64</td>
</tr>
<tr>
<td>6.3</td>
<td>Terrain Visualization</td>
<td>64</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Polygonal Information Visualization</td>
<td>64</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Combination with Terrain Texture</td>
<td>66</td>
</tr>
<tr>
<td>6.3.3</td>
<td>N-Best Views</td>
<td>70</td>
</tr>
<tr>
<td>6.3.4</td>
<td>Implementation Details</td>
<td>70</td>
</tr>
<tr>
<td>6.4</td>
<td>Conclusions</td>
<td>73</td>
</tr>
</tbody>
</table>

## 7 3D Shape Retrieval Using Viewpoint Measures

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1</td>
<td>Introduction</td>
<td>75</td>
</tr>
<tr>
<td>7.2</td>
<td>Background</td>
<td>76</td>
</tr>
<tr>
<td>7.2.1</td>
<td>Feature-Based Methods</td>
<td>77</td>
</tr>
<tr>
<td>7.2.2</td>
<td>Graph-Based Methods</td>
<td>77</td>
</tr>
<tr>
<td>7.2.3</td>
<td>View-Based Methods</td>
<td>77</td>
</tr>
<tr>
<td>7.3</td>
<td>View-Based Similarity Framework</td>
<td>78</td>
</tr>
<tr>
<td>7.3.1</td>
<td>$L_2$ Distance between Information Spheres</td>
<td>78</td>
</tr>
<tr>
<td>7.3.2</td>
<td>Earth Mover’s Distance between Information Histograms</td>
<td>80</td>
</tr>
<tr>
<td>7.3.3</td>
<td>Mutual Information Difference</td>
<td>81</td>
</tr>
<tr>
<td>7.4</td>
<td>Results and Discussion</td>
<td>81</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Experimental Results</td>
<td>81</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Discussion</td>
<td>86</td>
</tr>
<tr>
<td>7.5</td>
<td>Conclusions</td>
<td>88</td>
</tr>
</tbody>
</table>

## 8 Conclusions

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Contributions</td>
<td>89</td>
</tr>
<tr>
<td>8.2</td>
<td>Future Work</td>
<td>91</td>
</tr>
</tbody>
</table>

## Bibliography

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>93</td>
</tr>
</tbody>
</table>
Perceptual Information-Theoretic Measures for Viewpoint Selection and Object Recognition

Abstract:

Viewpoint selection has been an emerging area in computer graphics for some years, and it is now getting maturity with applications in fields such as scene navigation, volume visualization, object recognition, mesh simplification, and camera placement. But why is viewpoint selection important? For instance, automated viewpoint selection could play an important role when selecting a representative model by exploring a large 3D model database in as little time as possible. Such an application could show the model view that allows for ready recognition or understanding of the underlying 3D model. An ideal view should strive to capture the maximum information of the 3D model, such as its main characteristics, parts, functionalities, etc. The quality of this view could affect the number of models that the artist can explore in a certain period of time.

In this thesis, we present an information-theoretic framework for viewpoint selection and object recognition. From a visibility channel between a set of viewpoints and the polygons of a 3D model we obtain several viewpoint quality measures from the respective decompositions of mutual information. We also review and compare in a common framework the most relevant viewpoint quality measures for polygonal models presented in the literature.

From the information associated to the polygons of a model, we obtain several shading approaches to improve the object recognition and the shape perception. We also use this polygonal information to select the best views of a 3D model and to explore it. We use these polygonal information measures to enhance the visualization of a 3D terrain model generated from textured geometry coming from real data.

Finally, we analyze the application of the viewpoint quality measures presented in this thesis to compute the shape similarity between 3D polygonal models. The information of the set of viewpoints is seen as a shape descriptor of the model. Then, given two models, their similarity is obtained by performing a registration process between the corresponding set of viewpoints.
Mesures Perceptuals de Teoria de la Informació per a la Selecció de Punt de Vista i Reconeixement d'Objectes

Resum:

La selecció de punts de vista ha estat una àrea emergent en la computació gràfica des de fa alguns anys i ara està aconseguint la maduresa amb aplicacions en camps com la navegació d’una escena, la visualització de volums, el reconeixement d’objectes, la simplificació d’una malla i la col·locació de la càmera. Però per què és important la selecció del punt de vista? Per exemple, la automatització de la selecció de punts de vista podria tenir un paper important a l’hora de seleccionar un model representatiu mitjançant l’exploració d’una gran base de dades de models 3D en el menor temps possible. Aquesta aplicació podria mostrar la vista del model que permet el millor reconeixement o comprensió del model 3D. Un punt de vista ideal ha de captar la màxima informació del model 3D, com per exemple les seves principals característiques, parts, funcionalitats, etc. La qualitat d’aquest punt de vista pot afectar el nombre de models que l’artista pot explorar en un determinat període de temps.

En aquesta tesi, es presenta un marc de teoria de la informació per a la selecció de punts de vista i el reconeixement d’objectes. Obtenim diverses mesures de qualitat de punt de vista a través de la descomposició de la informació mútua d’un canal de visibilitat entre un conjunt de punts de vista i els polígons d’un model 3D. També revi-sem i comparem en un marc comú les mesures més rellevants que s’han presentat a la literatura sobre la qualitat d’un punt de vista d’un model poligonal.

A partir de la informació associada als polígons d’un model, obtenim diversos tipus de renderitzat per millorar el reconeixement d’objectes i la percepció de la forma. Utilitzem aquesta informació poligonal per seleccionar les millors vistes d’un model 3D i per la seva exploració. També usem aquestes mesures d’informació poligonal per millorar la visualització d’un model de terreny 3D amb textures generat a partir de dades reals.

Finalment, s’analitza l’aplicació de les mesures de qualitat de punt de vista presentades en aquesta tesi per calcular la similitud entre dos models poligonals. La informació del conjunt de punts de vista és vista com un descriptor del model. Llavors, donats dos models poligonals, la seva similitud s’obté mitjançant la realització d’un procés de registre entre els conjunts de punts de vista corresponents.
Medidas Perceptuales de Teoría de la Información para la Selección de Puntos de Vista y Reconocimiento de Objetos

Resumen:
La selección de puntos de vista ha sido un área emergente en la computación gráfica desde hace algunos años y ahora está alcanzando la madurez con aplicaciones en campos como la navegación de una escena, la visualización de volúmenes, el reconocimiento de objetos, la simplificación de una malla y la colocación de la cámara. Pero por qué es importante la selección de un punto de vista? Por ejemplo, la automatización de la selección de puntos de vista podría tener un papel importante a la hora de seleccionar un modelo representativo mediante la exploración de una gran base de datos de modelos 3D en el menor tiempo posible. Esta aplicación podría mostrar la vista del modelo que permite el mejor reconocimiento o comprensión del modelo 3D. Un punto de vista ideal debe captar la máxima información del modelo, como por ejemplo sus principales características, partes, funcionalidades, etc. La calidad de este punto de vista puede afectar el número de modelos que el artista puede explorar en un determinado periodo de tiempo.

En esta tesis, se presenta un marco de teoría de la información para la selección de puntos de vista y el reconocimiento de objetos. Obtenemos diversas medidas de calidad de punto de vista a través de la descomposición de la información mutua de un canal de visibilidad entre un conjunto de puntos de vista y los polígonos de un modelo 3D. También revisamos y comparamos en un marco común las medidas más relevantes que se han presentado en la literatura sobre la calidad de un punto de vista de un modelo poligonal.

A partir de la información asociada a los polígonos de un modelo, obtenemos varios tipos de renderizado para mejorar el reconocimiento de objetos y la percepción de la forma. Utilizamos esta información poligonal para seleccionar las mejores vistas de un modelo 3D y para su exploración. También usamos estas medidas de información poligonal para mejorar la visualización de un modelo de terreno 3D con texturas generado a partir de datos reales.

Finalmente, se analiza la aplicación de las medidas de calidad de punto de vista presentadas en esta tesis para calcular la similitud entre dos modelos poligonales. La información del conjunto de puntos de vista es considerada como un descriptor del modelo. Entonces, dados dos modelos poligonales, su similitud se obtiene mediante la realización de un proceso de registro entre los conjuntos de puntos de vista correspondientes.
CHAPTER 1

Introduction

1.1 Motivation

A 3D model is a digital representation of a real or imaginary object and a key feature in the Information Age. A large number of 3D models are used daily across diverse fields such as computer games, computer-aided design, interior design, visualization, simulation, and film industry. These 3D models can be computer-generated or done by artists or a 3D scanner and can be represented in different ways, such as voxels, polygons, point clouds, and nurbs (see Figure 1.1). When one of these 3D models needs to be visualized in a computer, two choices have to be made.

First, we have to decide which point of view of the object we should present. This could be a viewpoint where we can see a large part of the model or a great number of details. It could also be a view that we are used to or a view that is highly aesthetic. Second, we need to decide how we paint the object in order to perceive the shape as well as possible. One way would be to visualize the object in a photorealistic way but this would be expensive in terms of computation time and we should also decide the situation of the lights to illuminate the model. In both situations we should provide the user with as much information as possible in order to understand or recognize the object.

In this thesis we analyze different measures based on information theory to quantify the quality of a viewpoint, to represent a 3D polygonal model in different ways from the quantification of the polygonal information, and to find the similarity between different 3D models.

1.2 Objectives

The main goal of this thesis is to find good information-theoretic measures to improve the perception of 3D polygonal models and their recognition.

To reach this objective we aim to
Analyze the use of different decompositions of the mutual information of an information channel between a set of viewpoints and a set of polygons to quantify the quality of a viewpoint.

Analyze the performance of the most significant viewpoint quality measures presented in the literature and group all of them together in a common framework.

Quantify in different ways the information associated to a polygon from a 3D model and use this information for visualization, viewpoint selection, and object exploration.

Analyze the use of viewpoint quality measures to measure the similarity between two 3D models.

### 1.3 Thesis Outline

This dissertation is organized in eight chapters. Following this introduction, the next seven chapters are:

- **Chapter 2: Background and Previous Work**
  
  In this chapter, we review the pioneering work on the understanding of human perception and the recognition process. We also review the basic concepts of information theory since it is the mathematical basis of most of our contributions. Finally, the visibility channel between a set of viewpoints and a polygonal model is reviewed.
1.3. Thesis Outline

- **Chapter 3: Viewpoint Information**
  In this chapter, we present a new perspective to quantify the information associated with a viewpoint. The starting point is twofold: a visibility channel between a set of viewpoints and the polygons of an object, and two specific information measures introduced in the field of neuroscience to evaluate the significance of stimuli and responses in the neural code. These information measures are applied to the visibility channel in order to quantify the information associated with each viewpoint. A number of experiments show the performance of the proposed measures in best view selection.


- **Chapter 4: Survey of Viewpoint Selection Measures for Polygonal Models**
  In this chapter, we review and compare a significant amount of measures to select good views of a polygonal 3D model. These measures are classified in four categories and their performance is analyzed using a benchmark where several human subjects were asked to select the best view of different 3D models. We also review several fields where the viewpoint selection measures have been applied. All the viewpoint selection measures compared are implemented in a publicly available framework.

  The content of this chapter, titled *A survey of viewpoint selection methods for polygonal models*, has been submitted to ACM Transactions on Applied Perception.

- **Chapter 5: Information Measures for Object Understanding**
  In this chapter, we present a new information-theoretic framework for object understanding. Three specific information measures introduced in the field of neural systems are used to visualize the information associated with an object. We also present several ways of evaluating the shape information from the observer’s point of view. To do this, the polygonal information is ‘projected’ onto the viewpoints to quantify the information associated with a viewpoint and is used to select the $N$ best views and to explore the object. A number of experiments show the behavior of all proposed measures.

  The content of this chapter, titled *Information measures for object understanding*, has been published in Signal, Image and Video Processing, vol. 7, no. 3, pages 467–478, May 2013 [Bonaventura 2013a].

- **Chapter 6: Information Measures for Terrain Visualization**
  In this chapter, we apply the information-theoretic framework for object understanding presented in Chapter 5 to terrain visualization and terrain view selection. The polygonal information measures are used in order to enhance the perception of the terrain shape with the combination of the original terrain texture.
These polygonal information measures are also used to select the N best views of a terrain.

The content of this chapter, titled *Information measures for terrain visualization*, has been submitted to Computer & Geosciences.

- **Chapter 7: 3D Shape Retrieval Using Viewpoint Measures**

In this chapter, we present an information-theoretic framework to compute the shape similarity between 3D polygonal models. Given a 3D model, an information channel between a sphere of viewpoints around the model and its polygonal mesh is defined to compute the specific information associated with each viewpoint. The obtained information sphere is seen as a shape descriptor of the model. Then, given two models, their similarity is obtained by performing a registration process between the corresponding information spheres. The distance between the information histograms is also defined as a coarse measure of similarity, as well as the scalar value given by the mutual information of the channel. The performance of all these measures is tested using the Princeton Shape Benchmark database.

The content of this chapter, titled *3D shape retrieval using viewpoint information-theoretic measures*, has been published in Computer Animations and Virtual Worlds, vol. 26, no. 2, pages 147–156, 2015 [Bonaventura 2015]. This journal publication is an extension of the paper *Viewpoint information-theoretic measures for 3D shape similarity* published in Proceedings of the 12th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry (VRCAI’13), pages 183–190, November 2013 [Bonaventura 2013b].

- **Chapter 8: Conclusions**

In this chapter, conclusions of the thesis and future work will be presented, along with a summary of the publications related with this thesis.
CHAPTER 2

Background and Previous Work

Contents

2.1 Introduction .................................................. 5
2.2 Visual Perception and Object Recognition ..................... 5
2.3 Information Theory ............................................. 7
  2.3.1 Entropy .................................................. 7
  2.3.2 Mutual Information ....................................... 9
  2.3.3 Decomposition of Mutual Information ..................... 10
  2.3.4 Jensen’s Inequality ....................................... 12
2.4 Visibility Channel ............................................. 13

2.1 Introduction

In this chapter, first, we review the pioneer work on visual perception and the different schools of thought on the recognition process (Section 2.2). Second, we present the most basic information theory concepts as well as three different ways of decomposing the mutual information between two random variables (Section 2.3). Finally, we review the visibility channel between a set of viewpoints and a 3D polygonal model (Section 2.4).

2.2 Visual Perception and Object Recognition

The human visual system is classically described [Peters 2000] either in terms of its ability to recognize familiar three-dimensional objects as structural representations of their comprising part-components [Biederman 1987], or as multiple-view descriptions [Koenderink 1979, Edelman 1992, Bülthoff 1995]. Biederman [Biederman 1987] proposed that familiar object recognition can be conceptualized as a computational process by which the projected retinal image of a three-dimensional object is segmented at regions of deep concavity to derive a reduced representation of its simple geometric components (e.g., blocks, cylinders, wedges, and cones) and their spatial relations. Nonetheless, many studies have since proved that the visual system demonstrates preferential behavioral and neuronal responses to particular object views [Bülthoff 1995, Tarr 1997, Logothetis 1995]. Indeed, recognition behavior continues to be highly selective for previously learned views even when highly unique object parts with little self-occlusion
are made available for discrimination [Tarr 1997]. Naturally, this raises the question of which view(s) ought to be represented for a given object, so as to support robust visual recognition. Palmer et al. [Palmer 1981] found that participants tend to agree on the canonical view (or the most representative image) of each familiar object that would facilitate its recognition. They are often off-axis views, such as a top-down three-quarter view, that arguably reveals the largest amount of surface area. In contrast, Harman et al. [Harman 1999] allowed participants to learn novel 3D objects (objects with reduced effects of familiarity and functionality [Blanz 1999]) by exploring them in virtual reality. They found that their participants spent time exploring “plan” views, namely views that were on-axis (or orthogonal) and parallel to the object’s structural axis. Perrett and Harries [Perrett 1988] and Perrett et al. [Perrett 1992] found a similar preference for “plan” views in tool-like as well as “novel” objects. The mixed evidence could be due to the fact that view-canonicity can be expressed by multiple factors [Blanz 1999]: goodness for recognition (a good view for recognition shows the most salient and significant features and it is stable with respect to small transformations, and it avoids a high number of occluded features), familiarity (recognition is influenced by the views that are encountered more frequently and during the initial learning), functionality (recognition is influenced by the views that are most relevant for how we interact with an object), and aesthetic criteria (preferred views can be influenced by geometric proportions).

Blanz et al. [Blanz 1999] investigated the preferred views of different participants in two different tasks. In the first task, the participants had to select a view for a brochure. In the second task, participants were told an object and had to imagine it, and then selected the view on a displayed similar 3D model that matched the corresponding mental representation. While in the first task participants tried to avoid accidental views, in the second task users frequently selected frontal- or side-views. Blanz et al. suggest that this discrepancy can be due to the fact that mental images are subjected to internal storage and processing economy while in the photography task the participants try to select the view with as much information as possible.

Foster and Gilson [Foster 2002] studied the discrimination performance between sets of pairs of similar models at different orientation, to find out what was dependent on structure and what from viewpoint. The 3D models used in the extensive tests were formed by the concatenation, at variable angles, of cylinders with axes of variable curvature and length. They obtained that discrimination performance (measured by discrimination index) was the additive effect of viewpoint-invariant and structure-invariant performances, where the values of cue (or stimuli) considered (number, curvature and length of parts, and angle between them) could be factorized out in both of them. I.e., they established from their experiments that $d_i = [k_i + f(\theta)]\Delta c$, where $d_i$ was the discrimination index for cue $i$, $k_i$ a constant value depending on the cue considered, $f(\theta)$ is a function of the orientation angle $\theta$ and $\Delta c$ is the value of the cue (normalized for all cues). They thus reconciled the opposed views of Biederman [Biederman 1987] and Bülthoff [Bülthoff 1995] by integrating them into a single model.
2.3 Information Theory

In 1948, Claude Shannon published “A mathematical theory of communication” [Shannon 1948] which marks the beginning of information theory. In this paper, he defined measures such as entropy and mutual information, and introduced the fundamental laws of data compression and transmission. Information theory deals with the transmission, storage, and processing of information, and is used in fields such as physics, computer science, mathematics, statistics, economics, biology, linguistics, neurology, learning, image processing, and computer graphics.

In this section, we present some basic concepts of information theory. For more details, see the books by Cover and Thomas [Cover 1991], Yeung [Yeung 2008], and Sbert et al. [Sbert 2009].

2.3.1 Entropy

Let $X$ be a discrete random variable with alphabet $\mathcal{X}$ and probability distribution $\{p(x)\}$, where $p(x) = Pr\{X = x\}$ and $x \in \mathcal{X}$. In this thesis, $\{p(x)\}$ will be also denoted by $p(X)$ or simply $p$. This notation will be extended to two or more random variables.

The Shannon entropy $H(X)$ of a discrete random variable $X$ with values in the set $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$ is defined by

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x), \quad (2.1)$$

where $p(x) = Pr[X = x]$, the logarithms are taken in base 2 (entropy is expressed in bits), and we use the convention that $0 \log 0 = 0$, which is justified by continuity. We can use interchangeably the notation $H(X)$ or $H(p)$ for the entropy, where $p$ is the probability distribution $\{p_1, p_2, \ldots, p_n\}$. As $-\log p(x)$ represents the information associated with the result $x$, the entropy gives us the average information or uncertainty of a random variable. Uncertainty and information can be seen as opposite sides of the same coin. While entropy quantifies the uncertainty we have before an event, information is a measure of the reduction in that uncertainty after the event.

Some other relevant properties [Shannon 1948] of the entropy are

1. $0 \leq H(X) \leq \log n$

   - $H(X) = 0$ if and only if all the probabilities except one are zero, this one having the unit value, i.e., when we are certain of the outcome.
   - $H(X) = \log n$ when all the probabilities are equal. This is the most uncertain situation.

2. If we equalize the probabilities, entropy increases.

When $n = 2$, the binary entropy (Figure 2.1) is given by

$$H(X) = -p \log p - (1 - p) \log(1 - p), \quad (2.2)$$
Chapter 2. Background and Previous Work

where the variable $X$ is defined by

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}.$$  

If we consider another random variable $Y$ with probability distribution $p(y)$ corresponding to values in the set $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$, the joint entropy of $X$ and $Y$ is defined as

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in \mathcal{Y}} p(x, y) \log p(x, y), \quad (2.3)$$

where $p(x, y) = Pr[X = x, Y = y]$ is the joint probability.

The conditional entropy $H(Y|X)$ of a random variable $Y$ given a random variable $X$ is defined by

$$H(Y|X) = \sum_{x \in X} p(x) H(Y|x) \quad (2.4)$$

$$= \sum_{x \in X} p(x) \left( - \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \right) \quad (2.5)$$

$$= - \sum_{x \in X} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x), \quad (2.6)$$

where $p(y|x) = Pr[Y = y|X = x]$ is the conditional probability of $y$ given $x$ and $H(Y|x)$ is the entropy of $Y$ given $x$.

The Bayes theorem expresses the relation between the different probabilities:

$$p(x, y) = p(x)p(y|x) = p(y)p(x|y). \quad (2.7)$$

If $X$ and $Y$ are independent, then $p(x, y) = p(x)p(y)$.
The conditional entropy can be thought of in terms of a channel whose input is the random variable $X$ and whose output is the random variable $Y$. $H(X|Y)$ corresponds to the uncertainty in the channel input from the receiver’s point of view, and vice versa for $H(Y|X)$. Note that in general $H(X|Y) \neq H(Y|X)$.

The following properties are also fulfilled:

1. $H(X,Y) \leq H(X) + H(Y)$
2. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
3. $H(X) \geq H(X|Y) \geq 0$

### 2.3.2 Mutual Information

The mutual information $I(X;Y)$ between two random variables $X$ and $Y$ is defined by

$$I(X;Y) = H(X) - H(X|Y)$$

$$= H(Y) - H(Y|X)$$

$$= -\sum_{x \in X} p(x) \log p(x) + \sum_{y \in \mathcal{Y}} \sum_{x \in X} p(x,y) \log p(x|y)$$

$$= \sum_{x \in X} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log \frac{p(y|x)}{p(y)}$$

$$= \sum_{x \in X} \sum_{y \in \mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}.$$  \hspace{1cm} (2.12)

Mutual information represents the amount of information that one random variable, the output of the channel, gives (or contains) about a second random variable, the input of the channel, and vice versa, i.e., how much the knowledge of $X$ decreases the uncertainty of $Y$ and vice versa. Therefore, $I(X;Y)$ is a measure of the shared information between $X$ and $Y$.

Mutual information $I(X;Y)$ has the following properties:

1. $I(X;Y) \geq 0$ with equality if, and only if, $X$ and $Y$ are independent.

2. $I(X;Y) = I(Y;X)$
3. \( I(X; Y) = H(X) + H(Y) - H(X, Y) \)

4. \( I(X; Y) \leq H(X) \)

The relationship between all the above measures can be expressed by the Venn diagram, as shown in Figure 2.2.

The relative entropy or Kullback-Leibler distance \( D_{KL}(p, q) \) between two probability distributions \( p \) and \( q \) [Cover 1991, Yeung 2008], that are defined over the alphabet \( \mathcal{X} \), is given by

\[
D_{KL}(p, q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)},
\]

where, from continuity, we use the convention that \( 0 \log 0 = 0 \), \( a \log a = \infty \) if \( a > 0 \), and \( 0 \log 0 = 0 \).

The relative entropy is "a measure of the inefficiency of assuming that the distribution is \( q \) when the true distribution is \( p \)" [Cover 1991].

The relative entropy satisfies the information inequality \( D_{KL}(p||q) \geq 0 \), with equality only if \( p = q \). The relative entropy is also called discrimination and it is not strictly a distance, since it is not symmetric and does not satisfy the triangle inequality. Moreover, we have to emphasize that the mutual information can be expressed as

\[
I(X; Y) = D_{KL}(\{p(x, y)\}||\{p(x)p(y)\}).
\]  

### 2.3.3 Decomposition of Mutual Information

Given a communication channel \( X \rightarrow Y \), mutual information can be decomposed in different ways to obtain the information associated with a value (or symbol) in \( \mathcal{X} \) or \( \mathcal{Y} \). Next, we present different definitions of information that have been proposed in the field of neural systems to investigate the significance associated to stimuli and responses [Deweese 1999, Butts 2003].

For random variables \( S \) and \( R \), representing an ensemble of stimuli \( \mathcal{S} \) and a set of responses \( \mathcal{R} \), respectively, mutual information (see Equations 2.9 and 2.11) is given by

\[
I(S; R) = H(R) - H(R|S)
\]

\[
= H(R) - \sum_{s \in \mathcal{S}} p(s)H(R|s)
\]

\[
= \sum_{s \in \mathcal{S}} p(s) \sum_{r \in \mathcal{R}} p(r|s) \log \frac{p(r|s)}{p(r)},
\]

where \( p(r|s) \) is the conditional probability of value \( r \) given a known value \( s \), and \( p(S) = \{p(s)\} \) and \( p(R) = \{p(r)\} \) are the marginal probability distributions of the input and output variables of the channel, respectively. The capital letters \( S \) and \( R \) as arguments of \( p(.) \) or \( p(.,.) \) are used to denote probability distributions.

To quantify the information associated to each stimulus or response, \( I(S; R) \) can be
decomposed as

\[ I(S;R) = \sum_{s \in S} p(s) I(s;R) \]

(2.18)

\[ = \sum_{r \in R} p(r) I(S;r), \]

(2.19)

where \( I(s;R) \) and \( I(S;r) \) represent, respectively, the information associated to stimulus \( s \) and response \( r \). Thus, \( I(S;R) \) can be seen as a weighted average over individual contributions from particular stimuli or particular responses. The definition of the contribution \( I(s;R) \) or \( I(S;r) \) can be performed in multiple ways, but we present here the three most basic definitions denoted by \( I_1, I_2 \) [Deweese 1999], and \( I_3 \) [Butts 2003].

Given a stimulus \( s \), three specific information measures that fulfill Equation 2.18 are defined:

- The **surprise** \( I_1 \) can be directly derived from Equation 2.17, taking the contribution of a single stimulus to \( I(S;R) \):

\[ I_1(s;R) = \sum_{r \in R} p(r|s) \log \frac{p(r|s)}{p(r)}. \]

(2.20)

This measure expresses the surprise about \( R \) from observing \( s \). It can be shown that \( I_1 \) is the only positive decomposition of \( I(S;R) \) [Deweese 1999]. This positivity can be shown by observing that \( I_1(s;R) \) is the Kullback-Leibler distance [Cover 1991] between the conditional probability \( p(R|s) \) and the marginal distribution \( p(R) \).

- The **specific information** \( I_2 \) [Deweese 1999] can be derived from Equation 2.16, taking the contribution of a single stimulus \( s \) to \( I(S;R) \):

\[ I_2(s;R) = H(R) - H(R|s) \]

(2.21)

\[ = - \sum_{r \in R} p(r) \log p(r) + \sum_{r \in R} p(r|s) \log p(r|s). \]

The specific information \( I_2 \) of a particular response is defined as the reduction in uncertainty in the stimulus gained by the observation of that response [Butts 2003]. Thus, this measure expresses the change in uncertainty about \( R \) when \( s \) is observed. Note that \( I_2 \) can take negative values. This means that certain observations \( s \) do increase our uncertainty about the state of the variable \( R \).

- The **stimulus-specific information** \( I_3 \) is defined [Butts 2003] by

\[ I_3(s;R) = \sum_{r \in R} p(r|s) I_2(S;r) \]

(2.22)

and also fulfills Equation 2.18 (for a proof, see [Butts 2003]). The most informative (or significant) stimuli are those that cause the most informative responses. Thus, a large value of \( I_3(s;R) \) means that the states of \( R \) associated with \( s \) are very
informative in the sense of $I_2(S; r)$ (i.e., the specific information associated with response $r$). That is, the most informative input values $s$ are those that are related to the most informative output values $r$. Observe that $I_1(s; R)$ and $I_2(s; R)$ are obtained from both distributions $p(R)$ and $p(R | s)$, while $I_3(s; R)$ is a weighted sum of the measure $I_2(S; r)$, which is obtained from distributions $p(S)$ and $p(S | r)$.

Similar to the above definitions for a stimulus $s$, the information associated to a response $r$ could be defined. The properties of positivity and additivity of these measures have been studied in [Deweese 1999, Butts 2003]. A measure is additive when the information obtained about $S$ from two observations, $r_1 \in R_1$ and $r_2 \in R_2$, is equal to that obtained from $r_1$ plus that obtained from $r_2$ when $r_1$ is known. While $I_1$ is always positive and non-additive, $I_2$ can take negative values but is additive, and $I_3$ can take negative values and is non-additive. On the one hand, because of the additivity property, DeWeese and Meister [Deweese 1999] prefer $I_2$ against $I_1$ since they consider that additivity is a fundamental property of any information measure. On the other hand, Butts [Butts 2003] proposes some examples that show how $I_3$ identifies the most significant stimuli.

### 2.3.4 Jensen’s Inequality

Some important properties of information measures can be deduced from the Jensen’s inequality [Cover 1991].

A function $f(x)$ is convex over an interval $(a, b)$ (the graph of the function lies below any chord) if for every $x_1, x_2 \in (a, b)$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda) x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$  \hspace{1cm} (2.23)

A function is strictly convex if equality holds only if $\lambda = 0$ or $\lambda = 1$. A function $f(x)$ is concave (the graph of the function lies above any chord) if $-f(x)$ is convex.

For instance, $x \log x$ for $x \geq 0$ is a strictly convex function, and $\log x$ for $x \geq 0$ is a strictly concave function [Cover 1991].

**Jensen’s inequality:** If $f$ is convex on the range of a random variable $X$, then

$$f(E[X]) \leq E[f(X)],$$  \hspace{1cm} (2.24)

where $E$ denotes expectation. Moreover, if $f(x)$ is strictly convex, the equality implies that $X = E[X]$ with probability 1, i.e., $X$ is a deterministic random variable with $Pr[X = x_0] = 1$ for some $x_0$.

One of the most important consequences of Jensen’s inequality is the information inequality $D_{KL}(p||q) \geq 0$. Other previous properties can also be derived from this inequality.

Observe that if $f(x) = x^2$ (convex function), then $E[X^2] - (E[X])^2 \geq 0$. So, the variance is invariably positive.

If $f$ is substituted by the Shannon entropy, which is a concave function, we obtain
2.4. Visibility Channel

Several measures and concepts introduced in this thesis are based on a visibility channel built between a set of viewpoints and the polygons of a 3D model. From this channel we can quantify, for instance, the information associated with both a viewpoint and a polygon of a 3D model. Thus, in this section, we introduce the main elements of a visibility channel.

Feixas et al. [Feixas 2009] proposed a viewpoint selection framework from an information channel \( V \rightarrow Z \) between the random variables \( V \) (input) and \( Z \) (output), which represent, respectively, a set of viewpoints \( \mathcal{V} \) and the set of polygons \( \mathcal{Z} \) of an object. This channel is defined by a conditional probability matrix obtained from the projected areas of polygons at each viewpoint and can be interpreted as a visibility channel where the conditional probabilities represent the probability of seeing a determined polygon from a given viewpoint (Figure 2.3). Individual viewpoints are indexed by \( v \) and individual polygons by \( z \). The capital letters \( V \) and \( Z \) as arguments of \( p(\cdot) \) are used to denote probability distributions. For instance, while \( p(\mathcal{V}) \) denotes the probability of a single viewpoint \( v \), \( p(V) \) represents the input distribution of the set of viewpoints.

The three basic elements of the visibility channel are:

- **Conditional probability matrix** \( p(Z|V) \), where each element \( p(z|v) = \frac{a_z(v)}{a_z(v)} \) is defined by the normalized projected area of polygon \( z \) in the sphere of directions centered at viewpoint \( v \), where \( a_z(v) \) is the projected area of polygon \( z \) at viewpoint \( v \), and \( a_z(v) \) is the total projected area of all polygons in the sphere of directions. Conditional probabilities fulfill \( \sum_{z \in \mathcal{Z}} p(z|v) = 1 \).

- **Input distribution** \( p(V) \), representing the probability of selecting each viewpoint, is obtained from the normalization of the projected area of the object for each viewpoint. The input distribution can be interpreted as the importance assigned to each viewpoint \( v \).

where \( JS(\pi_1, \pi_2, \ldots, \pi_n; p_1, p_2, \ldots, p_n) \) is the Jensen-Shannon divergence of probability distributions \( p_1, p_2, \ldots, p_n \) with prior probabilities or weights \( \pi_1, \pi_2, \ldots, \pi_n \), fulfilling \( \sum_{i=1}^n \pi_i = 1 \). The JS-divergence measures how ‘far’ are the probabilities \( p_i \) from their likely joint source \( \sum_{i=1}^n \pi_i p_i \) and equals zero if and only if all the \( p_i \) are equal. It is important to note that the JS-divergence is identical to \( I(X;Y) \) when \( \pi_i = p(x_i) \) and \( p_i = p(Y|x_i) \) for each \( x_i \in \mathcal{X} \), where \( p(X) = \{p(x_i)\} \) is the input distribution, \( p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_m|x_i)\} \), \( n = |\mathcal{X}| \), and \( m = |\mathcal{Y}| \) [Burbea 1982, Slonim 2000].

The Jensen-Shannon inequality [Burbea 1982]:

\[
JS(\pi_1, \pi_2, \ldots, \pi_n; p_1, p_2, \ldots, p_n) \equiv H \left( \sum_{i=1}^n \pi_i p_i \right) - \sum_{i=1}^n \pi_i H(p_i) \geq 0, 
\]

where \( JS(\pi_1, \pi_2, \ldots, \pi_n; p_1, p_2, \ldots, p_n) \) is the Jensen-Shannon divergence of probability distributions \( p_1, p_2, \ldots, p_n \) with prior probabilities or weights \( \pi_1, \pi_2, \ldots, \pi_n \), fulfilling \( \sum_{i=1}^n \pi_i = 1 \). The JS-divergence measures how ‘far’ are the probabilities \( p_i \) from their likely joint source \( \sum_{i=1}^n \pi_i p_i \) and equals zero if and only if all the \( p_i \) are equal. It is important to note that the JS-divergence is identical to \( I(X;Y) \) when \( \pi_i = p(x_i) \) and \( p_i = p(Y|x_i) \) for each \( x_i \in \mathcal{X} \), where \( p(X) = \{p(x_i)\} \) is the input distribution, \( p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_m|x_i)\} \), \( n = |\mathcal{X}| \), and \( m = |\mathcal{Y}| \) [Burbea 1982, Slonim 2000].
Chapter 2. Background and Previous Work

(a) Viewpoint sphere

(b) Probability distributions of the visibility channel $V \rightarrow Z$

Figure 2.3: Visibility channel.

- Output distribution $p(Z)$, given by

$$p(z) = \sum_{v \in V} p(v)p(z|v), \quad (2.26)$$

which represents the average projected area of polygon $z$.

The mutual information of channel $V \rightarrow Z$, that expresses the degree of dependence or correlation between the set of viewpoints and the polygons of the model [Feixas 2009], is defined by

$$I(V;Z) = \sum_{v \in V} p(v) \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p(z)}. \quad (2.27)$$

From this visibility channel, different measures of viewpoint quality, such as viewpoint entropy [Vázquez 2001] and viewpoint mutual information [Feixas 2009], have been applied to viewpoint selection in computer graphics. These measures are reviewed in Chapter 3.
CHAPTER 3

Viewpoint Information

Contents

3.1 Introduction ............................................................... 15
3.2 Background ............................................................... 16
   3.2.1 Viewpoint Entropy ............................................... 16
   3.2.2 Viewpoint Kullback-Leibler .................................... 16
   3.2.3 Viewpoint Mutual Information ................................. 17
3.3 Viewpoint Information Measures .................................... 17
3.4 Results ................................................................. 19
3.5 Conclusions .............................................................. 22

3.1 Introduction

Why is viewpoint selection important? A large number of 3D models or objects are relied on daily across diverse fields such as computer game development, computer-aided design, and interior design. For instance, automated viewpoint selection could play an important role when an artist has to select a representative model by exploring a large 3D model database in as little time as possible. Such an application could show the model view that allows for ready recognition or understanding of the underlying 3D model. An ideal view should strive to capture the maximum information of the 3D model, such as its main characteristics, parts, functionalities, etc. The quality of this view could affect the number of models that the artist can explore in a certain period of time.

Best view selection is a fundamental task in object recognition and as we have seen in Section 2.2, many works have demonstrated that the recognition process is view-dependent [Palmer 1981, Tarr 1997, Blanz 1999]. In computer graphics, several viewpoint quality measures, such as viewpoint entropy and viewpoint mutual information, have been applied in areas such as best view selection for polygonal models [Vázquez 2001, Feixas 2009], scene exploration [Sokolov 2006], and volume visualization [Bordoloi 2005, Viola 2006].

In this chapter, we propose two new viewpoint quality measures that are respectively derived from two different decompositions of mutual information proposed by DeWeese and Meister [Deweese 1999] and Butts [Butts 2003] in the field of neural systems to quantify the information associated with stimuli and responses. First, we set
an information channel between a set of viewpoints and the polygons of an object and, then, we use those information measures to calculate the information associated with a viewpoint. Experimental results show the performance of these information measures to evaluate the quality of a viewpoint. This chapter is organized as follows. In Section 3.2, we present some previous work on viewpoint quality measures. In Section 3.3, two new viewpoint information measures are presented. In Section 3.4, experimental results show the behavior of the proposed measures to select the best views. Finally, in Section 3.5, our conclusions are presented.

3.2 Background

In this section, we present several information-theoretic viewpoint selection measures previously presented in the literature that lead us to the viewpoint quality measures introduced in this chapter.

3.2.1 Viewpoint Entropy

From Equation 2.1, Vázquez et al. [Vázquez 2001] defined the viewpoint entropy (VE) as the Shannon entropy of the probability distribution of the relative areas of the projected polygons over the sphere of directions centered at viewpoint $v$. Thus, the viewpoint entropy is defined by

$$VE_v = -\sum_{i=0}^{N_f} a_i \log \frac{a_i}{a_t},$$  \hspace{1cm} (3.1)

where $N_f$ is the number of polygons of the 3D polygonal model, $a_i$ is the projected area of polygon $i$ over the sphere, $a_0$ represents the projected area of background, and $a_t = \sum_{i=0}^{N_f} a_i$ is the total area of the sphere. The maximum entropy is obtained when a certain viewpoint can see all the polygons with the same projected area. The best viewpoint is defined as the one that has maximum entropy. In molecular visualization, both maximum and minimum entropy views show relevant characteristics of a molecule [Vázquez 2006].

3.2.2 Viewpoint Kullback-Leibler

From Equation 2.13, Sbert et al. [Sbert 2005] defined the viewpoint Kullback-Leibler distance (VKL) as

$$VKL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{A_t A_i},$$  \hspace{1cm} (3.2)

where $a_i$ is the projected area of polygon $i$, $a_t = \sum_{i=1}^{N_f} a_i$, $A_i$ is the actual area of polygon $i$ and $A_T = \sum_{i=1}^{N_f} A_i$ is the total area of the object. The VKL measure is interpreted as the distance between the normalized distribution of projected areas and the “ideal” projection, given by the normalized distribution of the actual areas. In this case, the projected area of the background can not be taken into account. The minimum value 0 is
obtained when the normalized distribution of projected areas is equal to the normalized distribution of actual areas. Thus, to select views of high quality means to minimize $V K L_v$.

### 3.2.3 Viewpoint Mutual Information

From the Equation 2.27, Feixas et al. [Feixas 2009] defined the viewpoint mutual information measure to select the most representative view of an object. The viewpoint mutual information of a viewpoint $v$ is defined by

$$VMI(v; Z) = \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p(z)}$$

(3.3)

and quantifies the degree of dependence between the viewpoint $v$ and the set of polygons. $VMI(v; Z)$ is interpreted as a measure of the quality of viewpoint $v$, where quality is considered here equivalent to representativeness.

The best viewpoint is defined as the one that has minimum VMI. High values of the measure mean a high dependence between viewpoint $v$ and the object, indicating a highly coupled view (for instance, between the viewpoint and a small number of polygons with low average visibility). On the other hand, the lowest values correspond to the most representative or relevant views, showing the maximum possible number of polygons in a balanced way.

It is important to observe that $VMI(v; Z) = D_{KL}(p(Z|v), p(Z))$, where $p(Z|v)$ is the conditional probability distribution between $v$ and the object and $p(Z)$ is the marginal probability distribution of $Z$, which in our case corresponds to the distribution of the average of projected areas. It is worth observing that $p(Z)$ plays the role of the target distribution in the $D_{KL}$ distance and also the role of the optimal distribution since the objective is that $p(Z|v)$ becomes similar to $p(Z)$ to obtain the best views. On the other hand, this role agrees with intuition since $p(Z)$ is the average visibility of polygon $z$ over all viewpoints, i.e., the mixed distribution of all views, and we can think of $p(Z)$ as representing, with a single distribution, the knowledge about the scene. Note that the difference between VMI (3.3) and VKL (3.2) is due to the fact that in the last case the distance is taken with respect to the actual areas. Viola et al. [Viola 2006] showed that the main advantage of VMI over VE is its robustness to deal with any type of discretisation or resolution of the volumetric dataset. The same advantage can be observed for polygonal data. Thus, while a highly refined mesh will attract the attention of VE, VMI will be almost insensitive to changes in the mesh resolution.

### 3.3 Viewpoint Information Measures

Inspired by the fact that VMI (Section 3.2.3) is obtained from a natural decomposition of mutual information, in this chapter we explore other mutual information decompositions of the visibility channel.
In this section, the information measures $I_1$, $I_2$ and $I_3$, derived from the decomposition of mutual information (Section 2.3.3), are applied to the visibility channel presented in Section 2.4. Although this perspective of analyzing the viewpoint quality is new, it is important to note that $I_1$ is equivalent to viewpoint mutual information (Section 3.2.3) and $I_2$ has a close relationship with viewpoint entropy (Section 3.2.1).

Given the visibility channel $V \rightarrow Z$, the viewpoint information is defined in the following three alternative ways:

- From (2.20), the viewpoint information $I_1$ of a viewpoint $v$ is defined as
  \[
  I_1(v;Z) = \sum_{z \in \mathcal{Z}} p(z|v) \log \frac{p(z|v)}{p(z)}. \tag{3.4}
  \]
  Observe that $I_1$ coincides with the viewpoint mutual information defined in [Feixas 2009] (see Equation 3.3). The lowest value of $I_1$ (i.e., $I_1(v;Z) = 0$) would be obtained when $p(Z|v) = p(Z)$.
  This means that the distribution of projected areas at a given viewpoint ($p(Z|v)$) would coincide with the average distribution of projected areas from all viewpoints ($p(Z)$). In this case, the view is considered maximally representative. Thus, while the most surprising views correspond to the highest $I_1$ values, the most representative ones correspond to the lowest $I_1$ values. The best viewpoint is defined as the one that has the lowest value of $I_1$ (i.e., maximum representativeness).

- From (2.21), the viewpoint information $I_2$ of a viewpoint $v$ is defined as
  \[
  I_2(v;Z) = H(Z) - H(Z|v) \tag{3.5}
  = \sum_{z \in \mathcal{Z}} p(z) \log p(z) + \sum_{z \in \mathcal{Z}} p(z|v) \log p(z|v).
  \]
  While the highest value of $I_2$ would correspond to a viewpoint that could only see one polygon, the lowest value of $I_2$ would be obtained if a viewpoint could see all polygons with the same projected area. In this case, the view is maximally diverse. The best viewpoint is defined as the one that has the lowest value of $I_2$ (i.e., maximum diversity).

Specific information $I_2(v;Z)$ is closely related to viewpoint entropy, defined as $H(Z|v)$ [Vázquez 2001, Feixas 2009], since $I_2(v;Z) = H(Z) - H(Z|v)$. As $H(Z)$ is constant for a given mesh resolution, $I_2(v;Z)$ and viewpoint entropy will essentially have the same performance in viewpoint selection because the highest value of $I_2(v;Z)$ corresponds to the lowest value of viewpoint entropy, and vice versa. An important drawback of viewpoint entropy is that it goes to infinity for finer and finer resolutions of the mesh (see [Feixas 2009]), while $I_2$ presents a more stable behavior due to the normalizing effect of $H(Z)$ in (3.5). The advantage of $I_2$ against viewpoint entropy could be appreciated in areas such as object recognition and mesh simplification. In the first case, the stable behavior of $I_2$ would enable us to compare the obtained values for objects with different mesh
3.4. Results

In this section, the behavior of $I_1$, $I_2$, and $I_3$ is analyzed. To calculate these measures, we need to obtain the projected area of every polygon for every viewpoint, and these areas will enable us to obtain the probabilities of the visibility channel ($p(V)$, $p(Z|V)$, and $p(Z)$). In this chapter, all measures have been computed without taking into account the background, and using a projection resolution of $640 \times 480$. In our experiments, all the objects are centered in a sphere of 642 viewpoints built from the recursive discretization of an icosahedron and the camera is looking at the center of this sphere. To obtain the viewpoint sphere, the smallest bounding sphere of the model is obtained and, then, the viewpoint sphere adopts the same center as the bounding sphere and a radius three times the radius of the bounding sphere.
Figure 3.1: (columns a, c, and e) The best view and (columns b, d, and f) the corresponding sphere of viewpoints of models (row i) lady of Elche, (row ii) coffee cup, (row iii) horse, and (row iv) ship, using (columns a–b) $I_1$, (columns c–d) $I_2$, and (columns e–f) $I_3$.

In Table 3.1 we show the number of polygons of the models used in this section and the cost of the preprocess step, i.e., the cost of computing the projected areas $a_z(v)$ and $a_t$. To show the behavior of the measures, the sphere of viewpoints is represented by a color map, where red and blue colors correspond respectively to the best and worst views. Remember that a good viewpoint corresponds to a low value of $I_1$ and $I_2$, and to high value of $I_3$. Our tests were run on an Intel Core i5 430M 2.27GHz machine with 4 GB RAM and an ATI Mobility Radeon HD 5470 with 512 MB.

To evaluate the performance of the viewpoint quality measures, four models have been used: a coffee cup, a horse, the Lady of Elche, and a ship. Figures 3.1 and 3.2 show, respectively, the best and worst views and the corresponding sphere of viewpoints for these models using measures (a–b) $I_1$, (c–d) $I_2$, and (e–f) $I_3$.

While the best views selected by $I_1$ show a global view of the object, the best views obtained by $I_2$ capture the maximum number of polygons in a balanced way (i.e., with a similar projected area). This means that $I_2$ has a high dependence of the resolution of the mesh, trying to see the areas with a finer discretization. On the contrary, it has been shown in [Feixas 2009] that $I_1$ is very robust with respect to the variation of the mesh resolution. The behavior of $I_3$ is very different of the one of $I_1$ and $I_2$ because the view with maximum $I_3$ tries to see the most informative polygons, that in general are placed...
3.4. Results

Figure 3.2: (columns a, c, and e) The worst view and (columns b, d, and f) the corresponding sphere of viewpoints of models (row i) lady of Elche, (row ii) coffee cup, (row iii) horse, and (row iv) ship, using (columns a–b) \( I_1 \), (columns c–d) \( I_2 \), and (columns e–f) \( I_3 \).

<table>
<thead>
<tr>
<th>Model</th>
<th># of polygons</th>
<th>Computational cost (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee cup</td>
<td>10732</td>
<td>3526</td>
</tr>
<tr>
<td>Horse</td>
<td>43571</td>
<td>3650</td>
</tr>
<tr>
<td>Ship</td>
<td>48811</td>
<td>3822</td>
</tr>
<tr>
<td>Lady of Elche</td>
<td>51978</td>
<td>3946</td>
</tr>
</tbody>
</table>

Table 3.1: Number of polygons of the models used and computational cost of the preprocessing step for each model in milliseconds.
in the most occluded, salient, and complex areas of the object. To better appreciate the behavior of $I_3$, the best and worst views (see column (e) in Figures 3.1 and 3.2) show the degree of informativeness of each polygon using a thermal scale, from blue (minimum information) to red (maximum information). Thus, it can be easily seen how $I_3$ selects the views with the highest informativeness. It is also important to note that a similar view can be considered as the best for one measure and the worst for another. See for instance the best and worst view of the coffee cup for $I_2$ and $I_1$, respectively (Figures 3.1(ii.c) and 3.2(ii.a)), and the best and worst view of the horse for $I_3$ and $I_1$, respectively (Figures 3.1(iii.e) and 3.2(iii.a)).

3.5 Conclusions

In this chapter, we have presented a new perspective based on the decomposition of mutual information to study the quality of a viewpoint. Two measures of specific information introduced in the field of neural systems have been adapted to quantify the information associated with a viewpoint ($I_2$ and $I_3$). These measures have been compared with viewpoint entropy and viewpoint mutual information, and several experiments have shown their performance in best view selection. The concepts of surprise, diversity, and informativeness associated with a viewpoint have been also discussed.
# Survey of Viewpoint Selection Measures for Polygonal Models

## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>23</td>
</tr>
<tr>
<td>4.2</td>
<td>Background</td>
<td>24</td>
</tr>
<tr>
<td>4.3</td>
<td>Viewpoint Selection Measures</td>
<td>26</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Notation</td>
<td>26</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Area Attributes</td>
<td>26</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Silhouette Attributes</td>
<td>32</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Depth and Stability</td>
<td>33</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Surface Curvature Attributes</td>
<td>35</td>
</tr>
<tr>
<td>4.4</td>
<td>Results, Discussion, and Applications</td>
<td>37</td>
</tr>
<tr>
<td>4.5</td>
<td>Conclusions</td>
<td>38</td>
</tr>
</tbody>
</table>

## 4.1 Introduction

The basic question underlying the viewpoint selection study and application is “what are good views of a 3D object or a scene?” In order to address this question, a number of computational measures have been proposed to quantify the goodness or the quality of a view. Depending on our goals, the best viewpoint can be, for instance, the view that allows us to see the largest number of parts of the object, the view that shows the most salient regions of the object, or the view that maximally changes when the underlying object is jittered. The main problem is how you decide if a viewpoint quality measure is better than another one. We can say when a view is good or not in an intuitive way but an impartial procedure is required to decide it. We need to set a benchmark where all the measures will be compared computing the best views for the same 3D models.

In this chapter, we review and compare a significant amount of measures to select good views of a polygonal 3D model. The computational measures reviewed are those that were motivated for “goodness for recognition” instead of other aspects such as familiarity and aesthetics. To compare these measures we have used the Dutagaci et al. benchmark [Dutagaci 2010] and they are classified according to recent work by Secord et al. [Secord 2011]. We also mention several fields where the viewpoint selection measures have been applied. The main contribution of this survey lies in collecting...
and testing the most basic measures introduced for viewpoint selection for polygonal models. We provide a publicly available framework where all the viewpoint selection measures compared are implemented.

This chapter is organized as follows. In Section 4.2, we review pioneering work in view-selection and the basic measures that have been proposed for estimating the quality of views. In Section 4.3, the most relevant measures are defined and described. In Section 4.4, we test the presented measures using the Dutagaci et al. benchmark [Dutagaci 2010]. We also present literatures that apply the viewpoint quality measures presented to other fields of research. Finally, in Section 4.5, our conclusions are presented.

4.2 Background

In this section we present the basis of viewpoint selection, that is, landmark research and the most basic measures that gave rise to other measures and methods that have been used in the last decade (Section 4.3). Section 4.4 also presents the application fields that these measures were employed in.

First, we review pioneer work on viewpoint selection. Attneave [Attneave 1954] analyzes informational aspects of visual perception and explains that information for object discrimination is concentrated along an object’s contour shape (i.e., 2D silhouette), especially where such information changes rapidly (i.e., peaks of curvature). Koenderink and van Doorn [Koenderink 1979] defined an aspect-graph, or visual potential of an object, where every node of the graph is an “aspect”, a data structure that contains the boundary information and singular points or paths visible from a given cell in the observer space. The edges in the graph represent visual events, i.e., transitions from cell to cell or change of aspect. Thus, this connected graph codifies any visual experience the observer can have when following an orbit around the object. Connolly [Connolly 1985] describes two algorithms that use partial octree models to determine the best next view to take. Kamada and Kawai [Kamada 1988] presented a measure to select a good view based on the angle between the view direction and the normal of the planes of the model. This method tries to avoid degenerative views, views where a plane is projected as a line and a line is projected as a point. Plemenos and Benayada [Plemenos 1996] extended Kamada’s work to insure that the user sees a great number of details. Plemenos’ measure takes into account the projected area and the number of polygons to evaluate the viewpoint goodness. Arbel and Ferrie [Arbel 1999] applied Shannon entropy to define entropy maps. These were used to guide an active observer along an optimal trajectory. Inspired by Kamada’s and Plemenos’ works, Vázquez et al. [Vázquez 2001] also used the Shannon entropy to quantify the information provided by a view. This measure incorporates both the projected area and the number of faces.

Weinshall and Werman [Weinshall 1997] define two measures: view likelihood and view stability. View likelihood measures the probability that a certain view of a given 3D object is observed and it may be used to identify “characteristic” views. View stability measures how little the image changes as the viewpoint is slightly perturbed and it may
be used to identify “generic” views. Stoev and Straßer [Stoev 2002] noticed that the projected area was not enough to visualize terrains and they presented a method that maximizes the maximum depth of the image in addition to the projected area. Given a sphere of viewpoints, Yamauchi et al. [Yamauchi 2006] computed the similarity between each two disjoint views using Zernike moments analysis and obtained a similarity weighted spherical graph. Here, a view was considered to be stable if all of the edges that were incident on its viewpoint in the spherical graph had high similarity weights.

Itti et al. [Itti 1998] maintain that visual attention is saliency-dependent and use a saliency map to represent the conspicuity or saliency at every location in the visual field by a scalar quantity. Thus, a good view could be described as one that is likely to be attended to, given its high saliency content. Borji and Itti [Borji 2013] have presented a state-of-the-art effort in visual attention modeling that can compute saliency maps from any image or video input. In [Lee 2005], mesh saliency is captured from surface curvatures. This is considered as a perception-inspired measure of regional importance and has been used in graphics applications such as mesh simplification and viewpoint selection. Gal and Cohen-Or [Gal 2006] introduced a method for partial matching of surfaces by using the abstraction of salient geometric features and a method to construct them.

Although they are not analyzed in this thesis, some measures that use semantic information of the model have also contributed to work on viewpoint selection. Thus, Secord et al. [Secord 2011], based on the work of Blanz et al. [Blanz 1999] and Gooch et al. [Gooch 2001], propose a measure that captures views from slightly above the horizon. Secord et al. [Secord 2011] also introduced a measure that tends to avoid views from directly below for objects that have an obvious orientation. The automatic method of Fu et al. [Fu 2008] can be used to determine both the base and the orientation of the object. When the model is a creature with eyes or a face, people prefer views where the eyes are visible [Zusne 1970]. Secord et al. [Secord 2011] have further proposed an attribute that simply sums all the visible pixels corresponding to the eyes’ surface. Finally, it is worth mentioning that Podolak et al. [Podolak 2006] have introduced a method to choose good viewpoints automatically by minimizing the symmetry of the object seen from the viewpoint.

Polonsky et al. [Polonsky 2005] and Secord et al. [Secord 2011] have described and analyzed a number of measures that were introduced to quantify the goodness of a view of an object. After analyzing different view descriptors, Polonsky et al. [Polonsky 2005] concluded that no single descriptor does a perfect job and have suggested that a combination of descriptors would amplify their respective advantage over each other. In this regard, Secord et al. [Secord 2011] have presented a perceptual model of viewpoint selection based on the combination of different attributes such as surface visibility, silhouette length, projected area, and maximum depth. If the region corresponding to the eyes’ surface is marked, Secord et al. have proposed changing the maximum depth according to eye preference. Dutagaci et al. [Dutagaci 2010] have presented a benchmark to validate best view selection methods. In this benchmark, 26 human subjects were asked to select the most informative view of 68 3D models. Dutagaci et al. [Dutagaci 2010] also provide a way to quantify the error of a best view selection algorithm.
compared to the data collected.

4.3 Viewpoint Selection Measures

In this section, the most relevant viewpoint selection measures are described according to several attributes captured from a particular viewpoint, such as area, silhouette, depth, stability, and surface curvature. For each measure, we give its definition and the reference of the paper where the measure was introduced. For the sake of completeness we have also included viewpoint entropy, viewpoint Kullback-Leibler, viewpoint mutual information, $I_1$, $I_2$, and $I_3$ already introduced in Chapter 3. Our classification has been inspired by the work of Secord et al. [Secord 2011]. All the measures of this section are tested in Section 4.4 and are available in a public common framework.

4.3.1 Notation

For comparison purposes between the analyzed measures, we propose a unified notation adopted from Feixas et al. [Feixas 2009], where an information channel was defined between the set of viewpoints $\mathcal{V}$ and the set of polygons $\mathcal{Z}$ (see Section 2.4). The projected area of polygon $z$ from viewpoint $v$ is denoted by $a_z(v)$ and the projected area of the model from viewpoint $v$ is given by $a_t(v)$. The quality of viewpoint $v$ will be expressed by $VQ(v)$.

Tables 4.1 and 4.2 show, respectively, the notation used in the measure definitions and the list of measures studied in this chapter. Observe that columns 3, 4, and 5 in Table 4.2 show the corresponding names used in surveys of Polonsky et al. [Polonsky 2005], Dutagaci et al. [Dutagaci 2010], and Secord et al. [Secord 2011], respectively. Column 6 indicates whether the best viewpoint corresponds to the highest (H) or the lowest (L) measure value. Column 7 shows whether the measure is sensitive (Y) to how the polygonal model is discretized or not (N). Finally, column 8 lists the main reference of the measure presented.

4.3.2 Area Attributes

The measures based on these attributes are computed using as main feature the area of polygons seen from a particular viewpoint.

**Number of visible triangles.** Plemenos and Benayada [Plemenos 1996] used the number of visible triangles seen from a viewpoint as a viewpoint quality measure. The higher the number of visible triangles the better the quality of a viewpoint. This measure is based on the fact that the most significant regions contain more details and, thus, more triangles. This measure is expressed as

$$VQ_1(v) = \sum_{z \in \mathcal{Z}} vis_z(v), \quad (4.1)$$
4.3. Viewpoint Selection Measures

- $z$: polygon
- $\mathcal{Z}$: set of polygons
- $v$: viewpoint
- $\mathcal{V}$: set of viewpoints
- $a_z(v)$: projected area of polygon $z$ from viewpoint $v$
- $a_t(v)$: projected area of the model from viewpoint $v$
- $\text{vis}_z(v)$: visibility of polygon $z$ from viewpoint $v$ (0 or 1)
- $N_p$: number of polygons
- $R$: number of pixels of the projected image
- $A_z$: area of polygon $z$
- $A_t$: total area of the model
- $p(z|v)$: conditional probability of $z$ given $v$
- $p(z)$: probability of $z$
- $p(v|z)$: conditional probability of $v$ given $z$
- $p(v)$: probability of $v$
- $H(V)$: entropy of the set of viewpoints
- $H(Z)$: entropy of the set of polygons
- $H(V|z)$: conditional entropy of the set of viewpoints given polygon $z$
- $H(Z|v)$: conditional entropy of the set of polygons given viewpoint $v$
- $\text{slength}(v)$: silhouette length from viewpoint $v$
- $\{h(\alpha)\}$: normalized silhouette curvature histogram
- $\alpha$: turning angle bin
- $\alpha$: turning angle between two consecutive pixels
- $\mathcal{A}$: set of turning angles
- $N_\alpha$: number of turning angles
- $\text{depth}(v)$: normalized maximum depth of the scene from viewpoint $v$
- $\{h(d)\}$: normalized histogram of depths
- $d$: depth bin
- $\mathcal{D}$: set of depth bins
- $N_d$: number of neighbors of $v$
- $L(v)$: size of the compression of the depth image corresponding to viewpoint $v$
- $L(v_i, v_j)$: size of the compression of the concatenation of the depth images corresponding to viewpoints $v_i$ and $v_j$
- $K_i$: curvature of vertex $i$
- $\{h(b)\}$: normalized histogram of visible curvatures from viewpoint $v$
- $b$: curvature bin
- $\mathcal{B}$: set of curvature bins
- $S(x)$: saliency of vertex $x$

Table 4.1: The most relevant notation symbols used in this chapter.
## Measure | Polonsky 2005 | Dutagaci 2010 | Secord 2011 | Ref
---|---|---|---|---
Visible triangles | H | Y | | [Plemenos 1996]
Projected area | View area | | | [Plemenos 1996]
Plemenos & Benayada | H | Y | | [Plemenos 1996]
Visibility ratio | | | | [Plemenos & Benayada 2005]
Visibility entropy | Surface area entropy | Surface area entropy | | [Vázquez 2001]
I2 | L | N | | [Bonaventura 2011]
Silhouette curvature | Silhouette curvature extrema | | | [Polonsky 2005]
Silhouette curvature | Silhouette curvature | | | [Bonaventura 2011]
Silhouette curvature | | | | [Polonsky 2005]
Silhouette curvature extrema | | | | [Bonaventura 2011]
Silhouette length | Idem | Idem | Idem | [Polonsky 2005]
Silhouette entropy | Idem | Idem | Idem | [Polonsky 2005]
Silhouette curvature extrema | | | | [Secord 2011]
Max depth | Max depth | Max depth | | [Vázquez 2009]
Max depth | | | | [Stoev 2002]
Viewpoint entropy | Surface area entropy | Surface area entropy | | [Vázquez 2009]
Maximum depth | Max depth | Max depth | | [Stoev 2002]
Viewpoint entropy | Surface area entropy | Surface area entropy | | [Vázquez 2009]
Mesh saliency | Mesh saliency | Mesh saliency | | [Lee 2005]
Projected saliency | Projected saliency | Projected saliency | | [Feixas 2009]
Saliency-based EVM | | | | [Feixas 2009]
Saliency-based visual stability | Idem | Idem | Idem | [Feixas 2009]

*Table 4.2: List of measures (column 1 and 2) with the corresponding names (columns 3, 4, and 5) used in surveys of Polonsky et al. [Polonsky 2005], Dutagaci et al. [Dutagaci 2010], and Secord et al. [Secord 2011], respectively. Column 6 indicates whether the best viewpoint corresponds to the highest (H) or the lowest (L) measure value. Column 7 shows whether the measure is sensitive (Y) to the polygonal discretization or not (N). Column 8 gives the main reference of the measure.*
where \( \text{vis}_z(v) \) is 1 if the polygon \( z \) is visible from viewpoint \( v \) and 0 otherwise. Different criteria can be used to consider whether a polygon is visible. In our implementation, a polygon is considered visible if at least one pixel of polygon \( z \) is visible from viewpoint \( v \) (\( a_z(v) > 0 \)). Obviously, the number of visible triangles is sensitive to the discretization of the model.

**Projected area.** Plemenos and Benayada [Plemenos 1996] also studied the projected area of the model from a viewpoint as a measure of viewpoint goodness since the number of visible triangles was found not to be enough in some cases. For example, if we consider a pencil, it is normal to have a high number of polygons around the pencil point. If we use the number of visible triangles to select the best viewpoint, we would only see a small part of the object. The projected area expressed as

\[
VQ_2(v) = a_t(v) \tag{4.2}
\]

can be considered as a viewpoint quality measure. Thus, the higher the projected area the better the viewpoint quality. This measure is insensitive to the discretization of the model.

**Plemenos & Benayada.** Plemenos and Benayada [Plemenos 1996] combined the number of visible triangles and the projected area to create a measure for viewpoint quality. A viewpoint is considered good if the percentage of the number of visible polygons plus the percentage of projected area with respect to the size of the screen is high. This measure can be expressed as

\[
VQ_3(v) = \frac{\sum_{z \in Z} \text{vis}_z(v) a_z(v)}{N} + \frac{\sum_{z \in Z} a_z(v)}{R}, \tag{4.3}
\]

where \( R \) is the total number of pixels of the image and \( N \) the total number of polygons (i.e., \( N = |Z| \)). For more details see also Barral et al. [Barral 1999]. Note that the first term is the ratio of visible polygons, where \( \frac{\text{vis}_z(v) a_z(v)}{a_z(v)+1} \) is equivalent to \( \text{vis}_z(v) \), and the second term is the ratio of the projected area with respect to the resolution of the screen. Thus, \( VQ_3(v) \) can be rewritten as

\[
VQ_3(v) = \frac{VQ_1(v)}{N} + \frac{VQ_2(v)}{R}. \tag{4.4}
\]

This measure is sensitive to polygonal discretization because \( VQ_1(v) \) is, as we have seen above.

**Visibility ratio.** Plemenos and Benayada [Plemenos 1996] also introduced the ratio between the visible surface area of the model from viewpoint \( v \) and the total surface area as a viewpoint quality measure. The visibility ratio is expressed by

\[
VQ_4(v) = \frac{\sum_{z \in Z} \text{vis}_z(v) A_z}{A_t}, \tag{4.5}
\]
where $A_z$ is the area of polygon $z$, and $A_t$ is the total area of the model. Observe that $A_z$ does not depend on the viewpoint because denotes the real area of polygon $z$. The best viewpoint corresponds to the minimum value of the measure. This measure is insensitive to the discretization of the model.

**Viewpoint entropy.** Vázquez et al. [Vázquez 2001, Vázquez 2003a] presented a measure for viewpoint selection based on Shannon entropy [Cover 1991, Yeung 2008]. This measure takes into account the projected area and the number of viewpoints and can be understood as the amount of information captured by a specific viewpoint. The viewpoint entropy is defined by

$$VQ_5(v) = H(v) = -\sum_{z \in Z} \frac{a_z(v)}{a_t(v)} \log \frac{a_z(v)}{a_t(v)}.$$ (4.6)

Using the notation of the visibility channel introduced in Section 2.4, the viewpoint entropy is rewritten as

$$VQ_5(v) = H(Z|v) = -\sum_{z \in Z} p(z|v) \log p(z|v),$$ (4.7)

where $H(Z|v)$ represents the conditional entropy of $Z$ given a viewpoint $v$. The best viewpoint corresponds to the one with maximum entropy which is obtained when a certain viewpoint can see all the faces with the same relative projected area. Viewpoint entropy is sensitive to polygonal discretization as in general the entropy increases with the number of polygons.

Polonsky et al. [Polonsky 2005] propose the application of viewpoint entropy using the probability of semantically important segments of the model.

**Information $I_2$.** Deweese and Meister [Deweese 1999] used a decomposition of mutual information in the field of neuroscience to quantify the information associated with stimuli and responses. Bonaventura et al. [Bonaventura 2011] applied this measure to the field of best viewpoint selection to express the informativeness of a viewpoint. The viewpoint information $I_2$ is defined by

$$VQ_6(v) = I_2(v; Z) = H(Z) - H(Z|v) = H(Z) - VQ_5(v) = -\sum_{z \in Z} p(z) \log p(z) + \sum_{z \in Z} p(z|v) \log p(z|v).$$ (4.8)

where $H(Z)$ stands for the entropy of model triangles. Note that $I_2$ is closely related to viewpoint entropy, defined as $H(Z|v)$ [Vázquez 2001, Feixas 2009], since $I_2(v; Z) = H(Z) - H(Z|v)$. As $H(Z)$ is constant for a given mesh resolution, $I_2(v; Z)$ and viewpoint entropy have the same performance in viewpoint selection because the highest value of $I_2(v; Z)$ corresponds to the lowest value of viewpoint entropy, and vice versa. An important drawback of viewpoint entropy is that it goes to infinity for finer and finer resolutions of the mesh (see [Feixas 2009]), while $I_2$ presents a more stable behavior.
due to the normalizing effect of $H(Z)$ in Equation 4.8. The best viewpoint is given by the one that has minimum $I_2$. Similarly to viewpoint entropy, this measure is also sensitive to polygonal discretization.

**Viewpoint Kullback-Leibler distance (VKL).** Sbert et al. [Sbert 2005] presented a viewpoint quality measure given by the Kullback-Leibler distance between the normalized distribution of the projected areas of polygons from viewpoint $v$ and the normalized distribution of the real areas of polygons. The viewpoint Kullback-Leibler distance is given by

$$VQ_7(v) = \sum_{z \in Z} \frac{a_z(v)}{a_t(v)} \log \frac{a_z(v)}{a_t(v)} A_z A_t. \quad (4.9)$$

Observe that the minimum value, which corresponds to the best viewpoint, is obtained when the normalized distribution of projected areas is equal to the normalized distribution of real areas. Viewpoint Kullback-Leibler distance is near insensitive to polygonal discretization.

**Viewpoint mutual information (or $I_1$).** Feixas et al. [Feixas 2009] presented a measure, called viewpoint mutual information (VMI), which captures the degree of correlation between a viewpoint and the set of polygons. Bonaventura et al. [Bonaventura 2011] renamed this measure as $I_1$ because this is one of the decomposition forms of mutual information used to deal with stimuli and responses [Deweese 1999]. The viewpoint mutual information is defined by

$$VQ_8(v) = VMI(v) = I_1(v; Z) = \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p(z)}. \quad (4.10)$$

High values of the measure mean a high correlation between viewpoint $v$ and the object, indicating a highly coupled view (for instance, between the viewpoint and a small number of polygons with low average visibility). On the other hand, the lowest values correspond to the most representative or relevant views (i.e., best viewpoints), showing the maximum possible number of polygons in a balanced way. VMI is insensitive to the discretization of the model. For more information see also Viola et al. [Viola 2006].

**Information $I_3$.** Butts [Butts 2003] introduced a new decomposition form of mutual information, called $I_3$, to quantify the specific information associated with a stimulus. Bonaventura et al. [Bonaventura 2011] proposed $I_3$ as a viewpoint quality measure. The measure $I_3$ is defined by

$$VQ_9(v) = I_3(v; Z) = \sum_{z \in Z} p(z|v) I_2(V; z), \quad (4.11)$$

where $I_2(V; z)$ is the specific information of polygon $z$ given by

$$I_2(V; z) = H(V) - H(V|z) = - \sum_{v \in V} p(v) \log p(v) + \sum_{v \in V} p(v|z) \log p(v|z), \quad (4.12)$$
where \( p(v|z) = \frac{p(v)p(z|v)}{p(z)} \) (Bayes theorem). Note that \( H(V) \) and \( H(V|z) \) represent the entropy of the set of viewpoints and the conditional entropy of the set of viewpoints given polygon \( z \), respectively. A high value of \( I_3(v;Z) \) means that the polygons seen by \( v \) are very informative in the sense of \( I_2(V;z) \). The most informative viewpoints are considered as the best views and correspond to the viewpoints that see the highest number of maximally informative polygons. The measure \( I_3 \) is sensitive to polygonal discretization.

### 4.3.3 Silhouette Attributes

The measures based on these attributes are computed using the silhouette of the object seen from a particular viewpoint. All these measures are insensitive to the discretization of the model because the polygons are not directly used.

**Silhouette length.** Polonsky et al. [Polonsky 2005] presented the silhouette length of the projected model from a viewpoint \( v \) as a measure of viewpoint goodness. The silhouette length is expressed as

\[
VQ_{10}(v) = sI(v), \quad (4.13)
\]

where \( sIengh(v) \) stands for the silhouette length from \( v \). In our implementation, the silhouette length of the model is computed from the viewpoint \( v \) by counting the number of pixels that belong to the silhouette. If there are multiple contours, the pixels of all the contours are added. The goodness of a viewpoint is associated with the maximum silhouette length.

**Silhouette entropy.** Polonsky et al. [Polonsky 2005] introduced the entropy of the silhouette curvature distribution, proposed by Page et al. [Page 2003], as a measure of viewpoint goodness. In our implementation, the silhouette curvature histogram is computed from the turning angles between consecutive pixels belonging to the silhouette. The range of the curvature is between \(-\pi/2\) and \(\pi/2\) with a step of \(\pi/4\) due to the angles obtained between neighbor pixels. The silhouette entropy is defined by

\[
VQ_{11}(v) = -\sum_{a=-\pi/2}^{\pi/2} h(\alpha) \log h(\alpha), \quad (4.14)
\]

where \( \{h(\alpha)\} \) represents the normalized silhouette curvature histogram and \( \alpha \) is the turning angle bin. The best viewpoint is the one with the highest silhouette entropy.

**Silhouette curvature.** Vieira et al. [Vieira 2009] introduced the complexity of the silhouette defined as the total integral of its curvature. In our implementation, the silhouette curvature is computed as

\[
VQ_{12}(v) = \frac{\sum_{c \in \mathcal{C}} |c|}{N_c}, \quad (4.15)
\]
4.3. Viewpoint Selection Measures

where $c$ is the turning angle between two consecutive pixels, $\mathcal{C}$ is the set of turning angles, and $N_c$ is the number of turning angles, equal to the number of pixels of the silhouette. The best viewpoint is given by the one with the maximum value.

**Silhouette curvature extrema.** As a variation of the above silhouette curvature measure, Secord et al. [Secord 2011] introduced the silhouette curvature extrema to emphasize high curvatures on the silhouette. The silhouette curvature extrema is computed as

$$VQ_{13}(v) = \frac{\sum_{c \in \mathcal{C}} \left( \frac{c}{\pi} \right)^2}{N_c}. \quad (4.16)$$

Similarly to silhouette curvature, the higher the value the better the viewpoint.

4.3.4 Depth and Stability

The measures based on these attributes are computed using the depth of the model seen from a particular viewpoint.

**Stoev & Straßer.** Stoev and Straßer [Stoev 2002] noticed that the projected area was not enough to visualize terrains because usually the view with most projected area is the one from above. They presented a method for camera placement that maximizes the maximum depth of the image in addition to the projected area. This measure is defined by

$$VQ_{14}(v) = \alpha p(v) + \beta d(v) + \gamma (1 - |d(v) - p(v)|), \quad (4.17)$$

where $p(v)$ is the normalized projection area from viewpoint $v$ and $d(v)$ is the normalized maximum depth of the scene from viewpoint $v$. For general purposes the authors proposed the use of the following values: $\alpha = \beta = \frac{1}{3}$ and $\gamma = \frac{1}{2}$. The Stoev & Straßer measure used in our implementation is given by

$$VQ_{14}(v) = \frac{1}{3} p(v) + \frac{1}{3} d(v) + \frac{1}{3} (1 - |d(v) - p(v)|). \quad (4.18)$$

For terrain scenarios, Stoev and Straßer [Stoev 2002] considered $\alpha = \beta = \frac{1}{4}$ and $\gamma = \frac{1}{2}$. The best viewpoint is the one with the maximum value, maximizing the projected area and the maximum depth and minimizing the difference between the projected area and the maximum depth. This measure is insensitive to polygonal discretization because the projected area and the maximum depth are insensitive too.

**Maximum depth.** Secord et al. [Secord 2011] considered only the maximum depth, used in Stoev and Straßer [Stoev 2002], as a descriptor of viewpoint quality. The maximum depth is expressed as

$$VQ_{15}(v) = \text{depth}(v). \quad (4.19)$$

As we have seen above, the maximum depth is insensitive to polygonal discretization and the best viewpoint is considered as the one with the maximum value.
**Depth distribution.** Instead of using only the maximum depth from a viewpoint, Secord et al. [Secord 2011] proposed a measure that maximizes the visible range of depths. The depth distribution measure defined by

\[ VQ_{16}(v) = 1 - \sum_{d \in \mathcal{D}} h(d)^2 \tag{4.20} \]

tries to capture the maximum diversity of depths, where \( d \) represents a depth bin, \( \mathcal{D} \) is the set of depth bins, and \( \{h(d)\} \) the normalized histogram of depths. The best viewpoint corresponds to the maximum value of the measure. This measure is insensitive to the discretization of the model.

**Instability.** Feixas et al. [Feixas 2009] defined viewpoint instability from the notion of dissimilarity between two viewpoints, which is given by the Jensen-Shannon divergence [Burbea 1982] between their respective projected area distributions. The use of Jensen-Shannon as a measure of view similarity was proposed by Bordoloi and Shen [Bordoloi 2005] in the volume rendering field. The viewpoint instability of \( v \) is defined by

\[ VQ_{17}(v) = \frac{1}{N_n} \sum_{j=1}^{N_n} D(v, v_j), \tag{4.21} \]

where \( v_j \) is a neighbor of \( v \), \( N_n \) is the number of neighbors of \( v \), and

\[ D(v, v_j) = JS \left( \frac{p(v)}{p(v) + p(v_j)}, \frac{p(v_j)}{p(v) + p(v_j)} ; p(Z|v), p(Z|v_j) \right) \]

is the Jensen-Shannon divergence between the distributions \( p(Z|v) \) and \( p(Z|v_j) \) captured by \( v \) and \( v_j \) with weights \( \frac{p(v)}{p(v) + p(v_j)} \) and \( \frac{p(v_j)}{p(v) + p(v_j)} \), respectively. The best viewpoint is the one with the lowest instability. The instability measure is sensitive to the discretization of the model.

**Depth-based visual stability.** Vázquez [Vázquez 2009] introduced a method to compute the view stability from the depth images of all viewpoints. The degree of similarity between two viewpoints is given by the normalized compression distance (NCD) between two depth images:

\[ \text{similarity}(v_i, v_j) = \text{NCD}(v_i, v_j) = \frac{L(v_i) - \min \left\{ L(v_i), L(v_j) \right\}}{\max \left\{ L(v_i), L(v_j) \right\}}, \tag{4.22} \]

where \( L(v_i) \) and \( L(v_j) \) are, respectively, the sizes of the compression of the depth images corresponding to viewpoints \( v_i \) and \( v_j \), and \( L(v_i) \) is the size of the compression of the concatenation of the depth images corresponding to \( v_i \) and \( v_j \).

Two views are considered similar if their distance is less than a given threshold. Hence, the most stable view is given by the one that has the largest number of similar
4.3. Viewpoint Selection Measures

views. The depth-based visual stability is given by

\[ VQ_{18}(v) = \text{#similar views to } v. \] (4.23)

This measure is robust to the discretization of the model because an image-based method is used. However, it is highly sensitive to the threshold value. The best view corresponds to the most stable one.

4.3.5 Surface Curvature Attributes

The measures based on these attributes are computed using the surface curvature of the shape. Note that, in the last two measures, area attributes are also taken into account.

Curvature entropy. Polonsky et al. [Polonsky 2005] propose a measure that evaluates the entropy of the curvature distribution over the visible portion of surface from a given viewpoint. This measure is inspired by the entropy of the Gaussian curvature distribution defined by Page et al. [Page 2003]. The curvature of vertex \(i\) is defined by

\[ K_i = 2\pi - \sum_j \phi_j, \] (4.24)

where the angle \(\phi_j\) is the wedge subtended by the edges of a triangle whose corner is at the vertex \(i\). The curvature entropy of a viewpoint \(v\) is defined by

\[ VQ_{19}(v) = -\sum_{b \in \mathcal{B}} h(b) \log h(b), \] (4.25)

where \(b\) represents a curvature bin, \(\mathcal{B}\) is the set of curvature bins, and \(\{h(b)\}\) the normalized histogram of visible curvatures from viewpoint \(v\). The higher the value the better the viewpoint. Curvature entropy is sensitive to the discretization of the model.

Visible saliency. Lee et al. [Lee 2005] presented a measure to select the best viewpoint based on the amount of saliency seen from a viewpoint. The saliency used is presented by Lee et al. [Lee 2005] and it is computed for every vertex using the curvature presented by Taubin [Taubin 1995]. The visible saliency measure is the sum of all the saliences of the vertices seen from viewpoint \(v\) and is defined by

\[ VQ_{20}(v) = \sum_{x \in \mathcal{X}} S(x), \] (4.26)

where \(\mathcal{X}\) is the set of visible vertices and \(S(x)\) the saliency of vertex \(x\). The saliency of vertex \(x\) is defined by

\[ S(x) = |G(C(x), \sigma) - G(C(x), 2\sigma)|, \] (4.27)

where \(G(C(v), \sigma)\) is the Gaussian-weighted average of the mean curvature. For more details see [Lee 2005]. The higher the value the better the viewpoint. Visual saliency is sensitive to polygonal discretization since the summation is done for the visible vertices.
Similarly to Lee et al. [Lee 2005], Sokolov and Plemenos [Sokolov 2005] presented a viewpoint quality measure given the sum of curvatures captured by a viewpoint where the curvature is computed as in Equation 4.24.

**Projected saliency.** Inspired by the visual saliency [Lee 2005], Feixas et al. [Feixas 2009] presented a method to select the best view using the saliency of the polygons. This saliency is computed for every polygon using an information channel between polygons and viewpoints. The projected saliency is defined by

\[
VQ_{21}(v) = \sum_{z \in Z} S(z) p(v|z),
\]  

where \(S(z)\) is the saliency of polygon \(z\) computed as

\[
S(z) = \frac{1}{N_z} \sum_{j=1}^{N_z} D(z, z_j),
\]

where polygon \(z_j\) is a neighbor of polygon \(z\), \(N_z\) is the number of neighbors of \(z\), and

\[
D(z, z_j) = JS \left( \frac{p(z)}{p(z) + p(z_j)}, \frac{p(z_j)}{p(z) + p(z_j)} \right) \cdot p(V|z), p(V|z_j)
\]

is the Jensen-Shannon divergence (Equation 2.25) between the distributions \(p(V|z)\) and \(p(V|z_j)\) with weights \(\frac{p(z)}{p(z) + p(z_j)}\) and \(\frac{p(z_j)}{p(z) + p(z_j)}\), respectively. For more details, see [Feixas 2009]. The higher the value the better the viewpoint. The projected saliency is sensitive to the discretization of the model. Similarly, other polygonal information measures have been projected to the viewpoints to select a good view [Bonaventura 2013a].

**Saliency-based EVMI.** Feixas et al. [Feixas 2009] presented an extended version of viewpoint mutual information (EVMI) where the target distribution is weighted by and importance factor. The importance-based EVMI is defined by

\[
VQ_{22}(v) = \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p'(z)},
\]  

where \(p'(z)\) is given by

\[
p'(z) = \frac{p(z) i(z)}{\sum_{z \in Z} p(z) i(z)},
\]

where \(i(z)\) is the importance of polygon \(z\). The saliency-based EVMI is obtained when \(i(z) = S(z)\) [Feixas 2009]. Similarly to VMI, the best viewpoint corresponds to the minimum value. Saliency-based EVMI is sensitive to polygonal discretization because the saliency of a polygon is sensitive too.

Serin et al. [Serin 2013] presented a similar measure where \(i(z)\) is given by the surface curvature and \(p(z)\) (i.e., average projected area) is substituted by the total area of the polygon.
4.4 Results, Discussion, and Applications

In this section, we test and compare the measures that were presented in Sections 4.3.2 to 4.3.5. These measures are computed for every model without taking into account any semantic information, such as the object’s preferred orientation.

First, we specify the details of the implementation that was used to compute the presented viewpoint selection measures. Second, we illustrate for all the measures the best view of three different 3D models. Third, the Dutagaci et al. benchmark [Dutagaci 2010] is used to analyze the accuracy of these measures in comparison with the best views selected by 26 human subjects.

The presented measures, except the visual saliency measure, have been implemented in a common framework. For the visual saliency measure (VQ20), we have used the Dutagaci’s implementation [Dutagaci 2010]. This is the only measure that is not included in the framework.

To compute the projected area of a polygon (usually triangle) we use a projection resolution of 640 x 640 pixels. No back-face culling optimization is applied and the polygons are rendered from both sides. All the models are centered inside a sphere of 642 viewpoints built from the recursive discretization of an icosahedron, and the camera is looking at the center of this sphere. To obtain the viewpoint sphere, the smallest bounding sphere of the model is obtained, and then, the viewpoint sphere adopts the same center as the bounding sphere and a radius that is six times the radius of the bounding sphere. This proportion between the viewpoint sphere and bounding sphere reduces at a minimum perspective distortion. The view-frustum of the camera is adjusted to ensure that only the model and the minimum background (19.2°) is seen. For the results of the depth-based visual stability measure (VQ18), we use a projection resolution of 128 x 128 pixels to reduce the computation time. In this case, the threshold used to decide if two viewpoints are similar is 0.87. Our framework, including the source code, is available at https://github.com/limdor/quoniam. In this framework, the user can add and test new measures.

To show the goodness of the viewpoint quality measures, three 3D models of the Dutagaci benchmark were used: the Standford Armadillo (17296 triangles), a cow (23216 triangles), and the Standford dragon (26142 triangles). In Figure 4.1 the best views that were selected by 26 human subjects in the Dutagaci et al. benchmark [Dutagaci 2010] are shown. Note that viewpoint entropy and information $I_2$ are grouped together in Figure 4.2 and in the following reported results because their performance is exactly the same (see Equation 4.8 in Section 4.3.2).

Figure 4.3 (from column (a) to column (u)) shows the best view and the corresponding sphere of viewpoints that is obtained with the following viewpoint quality measures: (a) number of visible triangles, (b) projected area, (c) Plemenos & Benayada, (d) visibility ratio, (e) viewpoint entropy / $I_2$, (f) viewpoint Kullback-Leibler distance, (g) viewpoint mutual information (or $I_1$), (h) $I_3$, (i) silhouette length, (j) silhouette entropy, (k) silhouette curvature, (l) silhouette curvature extrema, (m) Stoev & Straßer, (n) maximum depth, (o) depth distribution, (p) instability, (q) depth-based visual stability, (r) curvature entropy, (s) visual saliency, (t) projected saliency, and (u) saliency-
Figure 4.1: Set of best views for (a–b) the armadillo, (c–d) the cow, and (e–f) the dragon selected by the 26 human subjects of the Dutagaci et al. benchmark [Dutagaci 2010].

based EVMI. Rows (i), (iii), and (v) show, respectively, the best views of the armadillo, the cow, and the dragon, and rows (ii), (iv), and (vi) show the corresponding viewpoint sphere from the selected viewpoint. The sphere of viewpoints is represented by a color map, where red and blue colors correspond, respectively, to the best and worst viewpoints in terms of the corresponding viewpoint quality measures. From the different distributions we can see the preferred and unfavored regions, the transition between them, and also the stability of the measure with respect to small viewpoint variations.

Dutagaci et al. [Dutagaci 2010] presented a benchmark to evaluate a set of viewpoint selection methods. This benchmark uses the most informative view of 68 models chosen by 26 human subjects. An error between 0 and 1 and the average for all the models can be computed using the benchmark. In Figure 4.2 we show the box plot ordered by median (4.2a) and the mean +/- the standard deviation ordered by mean (4.2b) of the error of the models for each method. Observe that if we rank the measures in terms of mean and median, the sets of the five best ones are the same: projected saliency [Feixas 2009], the number of visible triangles [Plemenos 1996], viewpoint entropy and $I_2$ [Vázquez 2001, Bonaventura 2011], curvature entropy [Polonsky 2005], and Plemenos & Benayada [Plemenos 1996].

For the sake of completeness, we present here literatures that apply the viewpoint quality measures presented in Section 4.3 to other fields of research.

In Table 4.3, for each reference we specify the measure(s) used or inspired by, and the field of application. The fields of application considered in Table 4.3 are scene exploration (SE), image-based modeling and rendering (IBMR), volume visualization (VV), flow visualization (FV), shape retrieval (SR), mesh simplification (MS), molecular visualization (MV), and camera placement (CP).

Note that, for self-containment, we only include here the work related to the most relevant measures, presented in Section 4.3. Note also that some of the measures in Table 4.3 might not fully match the ones introduced in Section 4.3, but they are as closely related as to be considered the same one.

### 4.5 Conclusions

In this chapter, we have reviewed the most relevant viewpoint quality measures, which were motivated to support good recognition of polygonal models. We have extended
Figure 4.2: Plot of the error for each method running the Dutagaci et al. benchmark [Dutagaci 2010] that checks 68 different models.
Figure 4.3: The best view and the corresponding sphere of viewpoints of three models using different methods: (a) \# visible triangles, (b) projected area, (c) Plemenos & Benayada, (d) visibility ratio, (e) viewpoint entropy / $I_2$, (f) viewpoint Kullback-Leibler, (g) viewpoint mutual information ($I_1$), (h) $I_3$, (i) silhouette length, (j) silhouette entropy, (k) silhouette curvature, (l) silhouette curvature extrema, (m) Stoev & Straßer, (n) maximum depth, (o) depth distribution, (p) instability, (q) depth-based visual stability, (r) curvature entropy, (s) visual saliency, (t) projected saliency, and (u) saliency-based EVMI.
### 4.5. Conclusions

<table>
<thead>
<tr>
<th>Reference</th>
<th>Measure</th>
<th>SE</th>
<th>IBMR</th>
<th>VV</th>
<th>FV</th>
<th>SR</th>
<th>MS</th>
<th>MV</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Massios 1998]</td>
<td>1</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Barral 1999]</td>
<td>3</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Barral 2000]</td>
<td>3,4</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Vázquez 2003b]</td>
<td>5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Andújar 2004]</td>
<td>5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Bordoloi 2005]</td>
<td>5,17</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Takahashi 2005]</td>
<td>5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ji 2006]</td>
<td>5,20</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Vázquez 2006]</td>
<td>5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Viola 2006]</td>
<td>8,22</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Castelló 2007]</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[González 2007]</td>
<td>8</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Castelló 2008a]</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Castelló 2008b]</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Feixas 2009]</td>
<td>8</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ozaki 2010]</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Ruiz 2010]</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Castelló 2011]</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Eitz 2012]</td>
<td>2,10,16</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Serin 2012]</td>
<td>5</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Tao 2013]</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Sarikaya 2014]</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Bonaventura 2015]</td>
<td>6,8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.3: (column 1) Reference of the paper. (column 2) Measure used or inspired in. (columns 3–9) Field of application: scene exploration (SE), image-based modeling and rendering (IBMR), volume visualization (VV), flow visualization (FV), shape retrieval (SR), mesh simplification (MS), molecular visualization (MV) and camera placement (CP).
a previous existing classification by Secord et al. [Secord 2011], and we have implemented and compared them in a single framework, so as to allow for a fair comparison. As ground truth, we have used the Dutagaci et al. [Dutagaci 2010] user evaluation database. Our public framework allows including easily any new measure for comparison, or use another database as ground-truth. From the results, we have short-listed five measures (projected saliency [Feixas 2009], the number of visible triangles [Plemenos 1996], viewpoint entropy and I2 [Vázquez 2001, Bonaventura 2011], curvature entropy [Polonsky 2005], and Plemenos & Benayada [Plemenos 1996]) that effectively represent the viewpoint preferences of the users. Finally, we have also presented the application fields that the different measures have been employed in, given that their utility could vary according to the purposes that they were designed for.
5.1 Introduction

How well can we perceive shape from shading under diffuse illumination? In diffuse
shading, the image intensity is related to the degree of self-occlusion, for instance, the
concavities in the surface correspond to a darker shade in the image. According to
Thompson et al. [Thompson 2011], very little is known about how the visual system
uses this type of shading to estimate shape, although it has been suggested that the
brain could use the heuristic that “dark is deep” [Langer 2000]. In other words, darker
intensities in the image tend to be deeper in concavities [Thompson 2011]. Diffuse
global illumination is usually approximated by a much cheaper non-physically realistic
technique, ambient occlusion or obscurances [Zhukov 1998, Landis 2002, Iones 2003,
Méndez-Feliu 2009], that gives a photorealistic appearance to objects with complex
geometries by providing the visual shading cues associated with self-occlusion.

In this chapter, we present a new information-theoretic framework that allows to
a human observer to analyze and visualize the information associated with an object.
This work is based on a visibility channel between the polygons of an object and a set
of viewpoints, and three specific information measures introduced in the field of neural systems [Deweese 1999, Butts 2003]. We extend some previous work on both viewpoint quality and polygonal information introduced by Feixas et al. [Feixas 2009], González et al. [González 2008], and Bonaventura et al. [Bonaventura 2011]. We adopt two different perspectives. On the one hand, shape information is given by several polygonal information measures that provide us with different forms of perceiving the object shape. On the other hand, we present different viewpoint quality measures, obtained from the projection of the polygonal information onto the viewpoints, and also two algorithms to select the $N$ best views and to explore the object, respectively. Different experiments show the behavior of all these measures and algorithms. The main contributions of this chapter are the introduction of specific information measures to quantify the polygonal information, the use of Tsallis mutual information to analyze the polygonal information depending on an entropic index, the definition of new viewpoint quality measures based on different forms of polygonal information, and new algorithms for $N$ best views and object exploration.

This chapter is organized as follows. In Section 5.2, we present two generalizations of mutual information and basic information about obscurrences and ambient occlusion. In Section 5.3, we introduce new polygonal information measures that will be visualized in comparison with other measures previously introduced. In Section 5.4, we present new viewpoint quality measures to select the best views and to explore the object. Finally, in Section 5.5, our conclusions are presented.

5.2 Background

In this section, we present a generalized version of Shannon entropy and two different ways of generalizing mutual information.

5.2.1 Tsallis Information Measures

Rényi [Rényi 1961] proposed a generalized entropy which recovers the Shannon entropy as a special case and Harvda and Charvát [Harvda 1967] introduced a new generalized definition of entropy which also includes the Shannon entropy as a particular case. Tsallis [Tsallis 1988] used the Harvda-Charvát entropy in order to generalize the Boltzmann entropy in statistical mechanics. The introduction of this entropy responds to the objective of generalizing the statistical mechanics to non-extensive systems. For the objectives of this thesis we review the so-called Harvda-Charvát-Tsallis entropy or, simply, Tsallis entropy.

The Harvda-Charvát-Tsallis entropy $H_\alpha(X)$ of a discrete random variable $X$ is defined by

$$H_\alpha(X) = k \frac{1 - \sum_{x \in X} p(x)\alpha}{\alpha - 1},$$

where $k$ is a positive constant (by default $k = 1$) and $\alpha \in \mathbb{R} - \{1\}$ is called entropic index.
This entropy recovers the Shannon entropy (calculated with natural logarithms) when $\alpha \to 1$ and fulfills the properties of non-negativity and concavity (for $\alpha > 0$). If $X$ and $Y$ are independent, then the Harvda-Charvát-Tsallis entropy fulfills the non-additivity property:

$$H_\alpha(X, Y) = H_\alpha(X) + H_\alpha(Y) + (1 - \alpha)H_\alpha(X)H_\alpha(Y), \quad (5.2)$$

where $H_\alpha(X, Y)$ is the Tsallis joint entropy. The Tsallis conditional entropy $H_\alpha(Y|X)$ is defined by

$$H_\alpha(Y|X) = \sum_{x \in X} p(x)^\alpha H_\alpha(Y|x)$$

$$= \sum_{x \in X} p(x)^\alpha \left(1 - \sum_{y \in Y} p(y|x)^\alpha\right)/\alpha - 1, \quad (5.3)$$

where $H_\alpha(Y|x)$ is the Tsallis entropy of $Y$ known $x$.

Similar to Equation 2.12, the Tsallis mutual information $MI_\alpha(X; Y)$ is defined [Taneja 1988, Tsallis 1998] by

$$MI_\alpha(X; Y) = \frac{1}{1 - \alpha} \left(1 - \sum_{x \in X} \sum_{y \in Y} p(x, y)^\alpha / p(y)^{\alpha-1} p(y)^{\alpha-1}\right). \quad (5.4)$$

Another way of generalizing mutual information is the so-called Tsallis mutual entropy $ME_\alpha(X; Y)$, that, similar to Equation 2.11, is defined [Furuichi 2006] by

$$ME_\alpha(X; Y) = H_\alpha(Y) - H_\alpha(Y|X)$$

$$= H_\alpha(Y) - \sum_{x \in X} p(x)^\alpha H_\alpha(Y|x)$$

$$= H_\alpha(Y) - \sum_{x \in X} p(x)^\alpha \left(\sum_{y \in Y} p(y|x) - p(y|x)^\alpha / \alpha - 1\right). \quad (5.5)$$

Furuichi [Furuichi 2006] defined Tsallis mutual entropy for $\alpha > 1$ to ensure non-negativity, but for the purposes of this thesis this assumption is not necessary. Observe that both measures, $MI_\alpha(X; Y)$ and $ME_\alpha(X; Y)$, recover the Shannon mutual information (calculated with natural logarithms) when $\alpha \to 1$ and are different for $\alpha \neq 1$.

5.2.2 Obscurances and Ambient Occlusion

Illumination in real world is very complex and the simulation of the effects is a complex task. Imagine an environment where the illumination is mostly diffuse as for example in open air in a cloudy day. The illumination of every object is the product of many interreflections but we can notice that the objects that are more hidden are seen as darker. These effects can be reproduced with global illumination techniques but the computational cost is high.
Zhukov et al. [Zhukov 1998] and Iones et al. [Iones 2003] presented an efficient technique, called obscurances, that achieved some features of global illumination techniques in a much more economic way. This technique is much more simple and much less costly than global illumination, as in this case it is necessary to simulate the interaction of light between all the objects. The effect of obscurances can be considered as a pure geometric property of every point of the scene and can be computed evaluating the occlusion of the point with the objects around it.

In 2002, Landis [Landis 2002] and Bredow [Bredow 2002] presented a technique based on simplified obscurances and named it ambient occlusion. For a survey see Mendez and Sbert [Méndez-Feliu 2009].

### 5.2.2.1 The Obscurances Illumination Model

Let us take a look how to compute the obscurances introduced by [Zhukov 1998]. The obscurances of a point $P$ is defined as:

$$W(P) = \frac{1}{\pi} \int_{\omega \in \Omega} \rho(d(P, \omega)) \cos \theta \, d\omega$$

(5.6)

where:

- $d(P, \omega)$ is the distance from $P$ to the first intersection point in $\omega$ direction.

- $\rho$ is a monotone increasing function with values between 0 and 1 defined for all positive values. The result is 0 with distance 0, a value between 0 and 1 with a distance from 0 to $d_{\text{max}}$ and 1 with distances greater than $d_{\text{max}}$ (Figure 5.1).

- $\theta$ is the angle between direction $\omega$ and the normal at point $P$.

The integral is over the hemisphere oriented according to the surface normal. Then the $W(P)$ takes values from 0 to 1 where 0 means totally occluded and 1 means completely open.

We can see that we only consider a limited environment around $P$ and beyond it we will not consider the occlusions. To control this limited environment we use the parameter $d_{\text{max}}$, depending on amount of shadow that we want. $d_{\text{max}}$ will be in concordance
5.3 View-Based Polygonal Information

Figure 5.2: Obscurances in videogames. By courtesy of Àlex Méndez.

with the relative sizes of the objects with respect to the scene and with the size of the scene itself.

Obscurances computes the indirect light of the scene, the inter-reflections between objects. The direct light of the scene can be computed separately with other common techniques. It is very important that the indirect light computed with obscurances (Figure 5.2) looks similar than indirect light computed with other global illumination techniques, especially in average intensity.

For this reason, we have to combine $W(P)$ in the following way to get the final indirect illumination:

$$I(P) = R(P) \times I_A \times W(P)$$

That is, the obscurances at point $P$ is multiplied by the diffuse reflectivity ($R(P)$) at the point and by an average ambient intensity of the whole scene ($I_A$).

$I_A$ is computed assuming that light energy is distributed uniformly and illuminates all the objects with the same intensity:

$$I_A = \frac{R_{ave}}{1 - R_{ave}} \times \frac{1}{A_{total}} \sum_{i=0}^{n} A_i \times E_i$$

where $A_i$ and $E_i$ are the area and the emittance of the patch $i$ respectively, $A_{total}$ is the area of all patches and $R_{ave}$ is the average reflectivity of all patches weighted by its area:

$$R_{ave} = \frac{1}{A_{total}} \sum_{i=0}^{n} A_i \times R_i$$

5.3 View-Based Polygonal Information

In this section, we define the polygonal information measures derived from the information measures presented in Section 2.3.3.
5.3.1 Shannon Polygonal Information

As we have seen in the Sections 2.4 and 3.3, the information associated with each viewpoint has been obtained from the definition of the channel between the sphere of viewpoints and the polygons of the object. We now want to obtain the information associated with each polygon. To illustrate this new approach, the reversed channel \( Z \rightarrow V \) is considered, where \( Z \) is the input and \( V \) the output.

The three basic elements of the channel \( Z \rightarrow V \) are the conditional probability matrix \( p(V|Z) \), the input distribution \( p(Z) \), and the output distribution \( p(V) \). Observe that, using the Bayes theorem, each element \( p(v|z) \) of the conditional probability matrix can be computed as

\[
p(v|z) = \frac{p(v)p(z|v)}{p(z)}.
\]

The distributions \( p(Z) \) and \( p(V) \) have been defined in Section 2.4.

To obtain the information associated with a polygon of the object, the information measures \( I_1, I_2, \) and \( I_3 \) (Section 2.3.3) are now rewritten in the context of the channel \( Z \rightarrow V \):

- From Equation 2.20, the polygonal information \( I_1 \) of a polygon \( z \) is defined as

\[
I_1(z; V) = \sum_{v \in V} p(v|z) \log \frac{p(v|z)}{p(v)}.
\]

It is important to remark that the polygonal information \( I_1 \) was previously introduced by Feixas et al. \[Feixas 2009\] with the name of polygonal mutual information. Observe that \( I_1(z; V) \) is a Kullback-Leibler distance (Equation 2.13) between \( p(V|z) \) and \( p(V) \). Thus, low values of \( I_1(z; V) \) correspond to polygons that “see” the maximum number of viewpoints with a probability distribution \( p(V|z) \) similar to \( p(V) \), while high values indicate low visibility.

- From Equation 2.21, the polygonal information \( I_2 \) of a polygon \( z \) is defined as

\[
I_2(z; V) = H(V) - H(V|z)
\]

\[
= -\sum_{v \in V} p(v) \log p(v) + \sum_{v \in V} p(v|z) \log p(v|z).
\]

Observe that low values of \( I_2(z; V) \) are achieved by entropic polygons, that is, polygons that “see” the maximum number of viewpoints in a uniform way (i.e., with a uniform probability distribution). On the other hand, when a polygon “sees” few viewpoints, its entropy is low and the value of \( I_2 \) is high.

- From Equation 2.22, the polygonal information \( I_3 \) of a polygon \( z \) is defined as

\[
I_3(z; V) = \sum_{v \in V} p(v|z) I_2(Z; v),
\]

\[5.12\]
where $I_2(Z; v)$ is the specific information of viewpoint $v$ (Section 3.3) given by

$$I_2(Z; v) = H(Z) - H(Z|v)$$

(5.13)

$$= -\sum_{z \in Z} p(z) \log p(z) + \sum_{z \in Z} p(z|v) \log p(z|v).$$

Note that low values of viewpoint information $I_2(Z; v)$ correspond to high values of viewpoint entropy $H(Z|v)$, and vice versa (Section 3.3). Thus, low values of $I_3(z; V)$ are obtained when the viewpoints “seen” by the polygon $z$ are not informative in the sense of $I_2(Z; v)$ (i.e., these viewpoints are highly entropic). The opposite happens with high values.

In Section 5.3.3, we will show the behavior of all these measures.

### 5.3.2 Tsallis Polygonal Information

In Section 5.2.1, two generalized versions of mutual information, $MI_\alpha$ and $ME_\alpha$ (Equations 5.4 and 5.5), have been derived from the Kullback-Leibler form of mutual information (Equation 2.12) and from the definition of mutual information as a difference of entropies (Equation 2.9), respectively. On the other hand, the polygonal information measures $I_1(z; V)$ and $I_2(z; V)$ have been also derived from Equations 2.12 and 2.9. In a similar way, we now derive two generalized versions of polygonal information measures $I_1$ and $I_2$ from $MI_\alpha$ and $ME_\alpha$, respectively:

- Similarly to Equation 2.18, Equation 5.4 can be rewritten as

  $$MI_\alpha(V; Z) = \sum_{z \in Z} p(z) \frac{1}{1-\alpha} \left(1 - \sum_{v \in V} \frac{p(v|z)^\alpha}{p(v)^{\alpha-1}}\right)$$

  (5.14)

  where

  $$I_{1\alpha}(z; V) = \frac{1}{1-\alpha} \left(1 - \sum_{v \in V} \frac{p(v|z)^\alpha}{p(v)^{\alpha-1}}\right)$$

  (5.15)

  is called Tsallis polygonal information $I_1$.

- Similarly to Equation 2.18, Equation 5.5 can be written as

  $$ME_\alpha(V; Z) = \sum_{z \in Z} p(z)^\alpha \left(\frac{H(V)_\alpha}{\sum_{z \in Z} p(z)^\alpha} - \sum_{v \in V} \frac{p(v|z)^\alpha - p(v|z)^\alpha}{\alpha - 1}\right)$$

  (5.16)

  where

  $$I_{2\alpha}(z; V) = \frac{H(V)_\alpha}{\sum_{z \in Z} p(z)^\alpha} - \sum_{v \in V} \frac{p(v|z) - p(v|z)^\alpha}{\alpha - 1}$$

  (5.17)
### Table 5.1: Number of polygons of the models used and computational cost of the preprocessing step for each model in milliseconds.

<table>
<thead>
<tr>
<th>Model</th>
<th># of polygons</th>
<th>Computational cost (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lady of Elche</td>
<td>51978</td>
<td>1576</td>
</tr>
<tr>
<td>Angel</td>
<td>11758</td>
<td>1404</td>
</tr>
<tr>
<td>Coffee cup</td>
<td>6376</td>
<td>1389</td>
</tr>
<tr>
<td>Mini</td>
<td>42910</td>
<td>1575</td>
</tr>
<tr>
<td>Ogre head</td>
<td>4336</td>
<td>1388</td>
</tr>
<tr>
<td>Horse</td>
<td>48811</td>
<td>1627</td>
</tr>
</tbody>
</table>

is called Tsallis polygonal information $I_2$.

It is important to note that while $I_{1a}(z; V)$ is weighted by $p(z)$ in Equation 5.14, $I_{2a}(z; V)$ is weighted by $p(z)^{\alpha}$ in Equation 5.16. These factors appear in a natural way when the dependence on $p(z)$ is removed from $I_{1a}(z; V)$ and $I_{2a}(z; V)$.

#### 5.3.3 Results

In this section we analyze the performance of polygonal measures $I_1$, $I_2$, and $I_3$ in comparison with obscurances, introduced by Zhukov et al. [Zhukov 1998]. We also show the results obtained with the Tsallis-based measures $I_{1a}$ and $I_{2a}$.

To calculate all these measures, we need to obtain the projected area of every polygon from each viewpoint. Then, these areas will enable us to obtain the probability distributions of the visibility channel ($p(V)$, $p(Z|V)$, and $p(Z)$). In this chapter, all measures have been computed without taking into account the background, and using a projection resolution of 640 × 640. In our experiments, all the objects are centered in a sphere of 642 viewpoints built from the recursive discretization of an icosahedron and the camera is looking at the center of this sphere. To obtain the viewpoint sphere, the smallest bounding sphere of the model is computed and, then, the viewpoint sphere adopts the same center as the bounding sphere and a radius three times the radius of the bounding sphere. Our tests were run on a Intel® Core™ i7-2600K 3.40GHz machine with 16 GB RAM and an ATI Radeon™ HD 6950 with 2048 MB. In Table 5.1, we show the number of polygons of the models used in this section and the cost of the preprocessing step, that is, the cost of computing the projected areas $a_z(v)$ and $a_t(v)$.

In Figure 5.3, we show the obscurances (Figure 5.3(a)) and the polygonal information $I_1$ (Figure 5.3(b)), $I_2$ (Figure 5.3(c)), and $I_3$ (Figure 5.3(d)) for the lady of Elche and the angel models. We compute obscurances casting rays from polygons and averaging the distances to the hit point weighted by a square root function, as done in [Méndez-Feliu 2009]. To obtain the images, $I_1$, $I_2$, and $I_3$ have been normalized between 0 and 1 and subtracted from 1. Thus, low values of $I_1$ and $I_2$, corresponding to non-occluded polygons, are represented by bright colors (i.e., values near to 1) in the grey-map, while high values, corresponding to occluded polygons, are represented by
5.3. View-Based Polygonal Information

Figure 5.3: Visualization of (a) obscurances, (b) polygonal information $I_1$, (c) polygonal information $I_2$, and (d) polygonal information $I_3$ for the lady of Elche and the angel models.

Figure 5.4: (a, b) Visualization of polygonal information $I_2$ for the lady of Elche and (c, d) polygonal information $I_3$ for the coffee cup. Images have been generated (b,d) with and (a,c) without polygonal interpolation, respectively.

dark colors (i.e., values near to 0) in the grey-map. On the other hand, low values of $I_3$, corresponding to polygons that are seen by viewpoints with low values of viewpoint information $I_2$, are represented by bright colors, and vice versa. Observe that the performance of $I_1$ (Figure 5.3(b)) and $I_2$ (Figure 5.3(c)) is very similar and can be interpreted as a kind of obscurances or ambient occlusion [Zhukov 1998, Landis 2002, Iones 2003]. On the other hand, using $I_3$ (see Figure 5.3(d)), we obtain a non-photorealistic visualization in the sense that it cannot not be obtained with a physically based rendering of the object, and that permits us to perceive the shape of the object in a novel and expressive way. Although we compute the information for each polygon, in the images of this chapter, the polygonal information is interpolated to obtain a smoother visualization. Figure 5.4 shows the difference of applying or not the polygonal interpolation to two different models (lady of Elche and coffee cup), where the grey-map has been obtained from the polygonal information measures $I_2$ and $I_3$, respectively.

In Figures 5.5 and 5.6, we show the Tsallis polygonal information measures $I_1$ and $I_2$ depending on the entropic index $\alpha$. From left to right, more contrasted images resulting in sharper shading are obtained with lower $\alpha$ values while the contrast almost vanishes with higher values. Note the similar behavior of $I_1$ and $I_2$ for all $\alpha$ values.
Figure 5.5: Grey-map representation of Tsallis polygonal information $I_1$ depending on the $\alpha$-value. Lower $\alpha$ values result in sharper shading.

Figure 5.6: Grey-map representation of Tsallis polygonal information $I_2$ depending on the $\alpha$-value. Lower $\alpha$ values result in sharper shading.
Figure 5.7: Visualization of (a) obscurances, (b) Tsallis polygonal information $I_1$ with $\alpha = 0.6$, and (c) Tsallis polygonal information $I_2$ with $\alpha = 0.6$. 
In Figure 5.7, the results obtained with Tsallis polygonal information $I_1$ and $I_2$ with $\alpha = 0.6$ are compared with obscurances. Observe that for some models as the car, the shading based on $I_1$ and $I_2$ can help us to spot better some details that are difficult to see with the obscurances. In Figure 5.8, we show the effect of combining a textured model with the Tsallis polygonal information $I_2$ with $\alpha = 0.6$.

5.4 Viewpoint Selection and Object Exploration

In this section, we present three new viewpoint quality measures based on the projection of the polygonal information onto the viewpoints, and we introduce two algorithms to select the $N$ best views and to explore an object, respectively.

5.4.1 Viewpoint Selection

First, we introduce several viewpoint quality measures based on the polygonal information measures introduced in Section 5.3. Then, the behavior of these measures is compared with the viewpoint information measures $I_1$ and $I_2$ (Section 3.3), which are respectively equivalent to the viewpoint mutual information and the complementary of viewpoint entropy [Feixas 2009, Vázquez 2001] (i.e., viewpoint information $I_2$ performs inversely to viewpoint entropy and, thus, the view with maximum entropy coincides with the view with minimum $I_2$, and vice versa).

The new measures of the quality of a viewpoint are obtained by projecting (or spreading) the polygonal information to the sphere of viewpoints. This method is similar to the one used by Feixas et al. [Feixas 2009] to obtain the saliency of a viewpoint. The projection of the polygonal information over a viewpoint $v$ is done by weighting the polygonal information of polygon $z$ by the transition probability $p(v|z)$ and summing over all polygons.

From the polygonal information measures $I_1$, $I_2$ and $I_3$, the viewpoint quality of $v$ is defined by

$$VQ_i(v) = \sum_{z \in \mathcal{Z}} p(v|z)I'_i(z; V),$$

where $i$ stands for the values 1, 2, or 3; $I'_1(z; V)$ and $I'_2(z; V)$ are given by $I_1(z; V)$ and $I_2(z; V)$ linearly scaled between 0 and 1; and $I'_3(z; V)$ is given by $1 - I_3(z; V)$ with
5.4. Viewpoint Selection and Object Exploration

I_3(z; V) linearly scaled between 0 and 1. Observe that high values of VQ will correspond to viewpoints that see the most complex parts of the model which are represented by the areas with more occlusions or significant details (i.e., with high values of polygonal information I_1 and I_2, and low values of polygonal information I_3). On the other hand, low values of VQ correspond to viewpoints that see the smoothest areas of the model, with small changes in visibility and less detail (i.e., with the lowest values of polygonal information I_1 and I_2, and the highest values of polygonal information I_3).

To evaluate the performance of these viewpoint quality measures, three models have been used: the Lady of Elche, the coffee cup, and the horse. Figures 5.9 and 5.10 has been organized as follows. From (a) to (e), we show the results of viewpoint information I_1 (called viewpoint mutual information), viewpoint information I_2 (which performs inversely to viewpoint entropy), VQ_1, VQ_2, and VQ_3, respectively. Figure 5.9 shows the “best” views and Figure 5.10 shows the “worst” views. Note that the models of (c-e) have been visualized with the polygonal information used to compute the corresponding measure VQ. It is important to note that the best views for the viewpoint information I_1 and I_2 correspond to the lowest values of these measures, and while the best view for I_1 shows a representative view of the object (from a geometrical perspective), the best view obtained by I_2 captures the maximum number of polygons in a uniform way (maximum entropy). Hence, I_2 is highly dependent on the resolution of the mesh, trying to see the areas with a finer discretization. On the other hand, the best views for the viewpoint quality measures VQ correspond with the highest values of these measures, showing the most complex parts of the object. These experiments show the good performance of the viewpoint quality measures VQ that capture the maximum information of the model coming from the areas with more details and saliency. Observe that, in the shown examples, I_2 obtains similar best views to VQ since usually views with maximum entropy see highly complex, very refined, areas. The contrary would happen for the worst views.
Figure 5.10: Worst views for (a) the viewpoint information $I_1$, (b) viewpoint information $I_2$, (c) $VQ_1$, (d) $VQ_2$, and (e) $VQ_3$.

5.4.2 N Best Views and Object Exploration

In order to understand or model an object, we are interested in selecting a set of representative views which provides an approximate representation of the object. With this goal in mind, a new viewpoint selection algorithm based on the viewpoint information $I_1$ extended with the polygonal information $I_2$ (or polygonal information $I_3$) is presented. Due to the similar behavior of the polygonal information measures $I_1$ and $I_2$, we only explore the performance of extending the viewpoint information $I_1$ with the polygonal information $I_2$ and $I_3$.

To compute the most representative set of views, Feixas et al. [Feixas 2009] proposed a viewpoint information $I_1$-based algorithm to select the set of viewpoints that minimize the $I_1(\bar{v}; Z)$ value of a set of views $\bar{v}$:

$$I_1(\bar{v}; Z) = \sum_{z \in Z} p(z|\bar{v}) \log \frac{p(z|\bar{v})}{p(z)}$$  \hspace{1cm} (5.19)

where

$$p(\bar{v}) = \sum_{v \in \bar{v}} p(v)$$  \hspace{1cm} (5.20)

and

$$p(z|\bar{v}) = \frac{\sum_{v \in \bar{v}} p(v)p(z|v)}{p(\bar{v})}.$$  \hspace{1cm} (5.21)

Due to the fact that this algorithm is NP-complete, a greedy solution was used. The viewpoint with minimum $I_1(v; Z)$ is selected as the first element of the set. Then, the next viewpoint selected is the one that minimizes $I_1(\bar{v}; Z)$, where $\bar{v}$ represents the virtual viewpoint that results from the clustering of the first two viewpoints. This process is repeated until $N$ views are selected.

We now define an extended viewpoint measure $EI_1$, where the target distribution
5.4. Viewpoint Selection and Object Exploration

$p(z)$ is weighted by an importance distribution based on the polygonal information measures $I_2$ or $I_3$. The measure $EI_1$ will be used to compute the $N$ best views. Observe that, using the polygonal information measures $I_2$ or $I_3$ as importance distribution, we prioritize to see the polygons with a high informativeness, that is, the ones with the most relevant shape information.

The extended viewpoint information $EI_1(v; Z)$ is defined by

$$EI_1(v; Z) = \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p'(z)},$$

where the target distribution $p'(z)$ is given by

$$p'(z) = \frac{p(z) imp(z)}{\sum_{z \in Z} p(z) imp(z)},$$

and the importance factor $imp(z)$ is given by $I'_2(z; V)$ or $I'_3(z; V)$ defined in Section 5.4.1.

Using the extended viewpoint information measure $EI_1(v; Z)$, our best view algorithm proceeds in the same way as the $I_1$-based algorithm [Feixas 2009] but we now minimize $EI_1(\hat{v}; Z)$ instead of minimizing $I_1(\hat{v}; Z)$. The minimization of $EI_1(\hat{v}; Z)$ is based on the fact that we are interested in minimizing the Kullback-Leibler distance between the distribution of projected areas captured by the viewpoints and the target distribution $p'(z)$ given by the average projected area of all polygons weighted by the importance distribution. In Figures 5.11 and 5.12, we show the results of selecting the six best views using the $I_1$-based algorithm without importance distribution [Feixas 2009] (column 1) and with importance distribution given by the polygonal information measures $I_2$ (column 2) and $I_3$ (column 3), respectively. The models of columns 2 and 3 have been rendered using polygonal information $I_2$ and polygonal information $I_3$, respectively, to illustrate how the algorithm selects the views depending on the polygonal information. Observe how our $EI_1$-based algorithm focuses the attention on the most informative parts of the model and enhances the view selection achieved with the $I_1$-based algorithm.

Finally, we present an exploratory algorithm, called exploratory tour, that first selects the best viewpoint (i.e., with minimum $EI_1(v; Z)$) and then successively visits the neighbor viewpoints that minimize the value of $EI_1(\hat{v}; Z)$ of all visited viewpoints. This algorithm is similar to the one presented by Feixas et al. [Feixas 2009] but now the selection of the successive viewpoints is also guided by the informativeness of polygons. Figures 5.13 and 5.14 show the performance of our $EI_1$-based algorithm using the polygonal information measures $I_2$ and $I_3$ as importance factors, respectively. The models of Figures 5.13 and 5.14 have been rendered using polygonal information $I_2$ and polygonal information $I_3$, respectively, to show how the exploration depends on the polygonal information. In these examples, the stopping criteria used by the exploration algorithms guided by the polygonal information $I_2$ and $I_3$ are given by the 25% and 15% of the initial value $EI_1(v; Z)$, respectively.
<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image 1" /></td>
<td><img src="image2.png" alt="Image 2" /></td>
<td><img src="image3.png" alt="Image 3" /></td>
</tr>
<tr>
<td><img src="image4.png" alt="Image 4" /></td>
<td><img src="image5.png" alt="Image 5" /></td>
<td><img src="image6.png" alt="Image 6" /></td>
</tr>
<tr>
<td><img src="image7.png" alt="Image 7" /></td>
<td><img src="image8.png" alt="Image 8" /></td>
<td><img src="image9.png" alt="Image 9" /></td>
</tr>
<tr>
<td><img src="image10.png" alt="Image 10" /></td>
<td><img src="image11.png" alt="Image 11" /></td>
<td><img src="image12.png" alt="Image 12" /></td>
</tr>
</tbody>
</table>

Figure 5.11: Six best views of the lady of Elche using (column 1) the $I_1$-based algorithm and the $EI_1$-based algorithm weighted by the polygonal information (column 2) $I_2$ and (column 3) $I_3$. 
Figure 5.12: Six best views of the angel using (column 1) the $I_1$-based algorithm and the $EI_1$-based algorithm weighted by the polygonal information (column 2) $I_2$ and (column 3) $I_3$. 
Figure 5.13: Exploratory tour with the extended viewpoint information $EI_1$ weighted by the polygonal information $I_2$.

Figure 5.14: Exploratory tour with the extended viewpoint information $EI_1$ weighted by the polygonal information $I_3$. 
5.5 Conclusions

In this chapter, we have presented an information theory framework for object understanding. From the definition of visibility channel between the polygons of an object and a set of viewpoints, we obtain several shading approaches using the polygonal information. Two of our shading measures (polygonal information $I_1$ and polygonal information $I_2$) are perceptually related to diffuse shading, where image intensity depends on the degree of self-occlusion, and a third measure (polygonal information $I_3$) represents a novel perceptual approach. Several results show that those measures improve on perceiving shape on similar ambient occlusion measures and that our viewpoint quality measures perform well in capturing object complexity. Finally, we apply the polygonal information measures to select best views and to explore an object.
CHAPTER 6

Information Measures for Terrain Visualization

Contents

6.1 Introduction ........................................ 63
6.2 Background ........................................ 64
6.3 Terrain Visualization .............................. 64
  6.3.1 Polygonal Information Visualization ........... 64
  6.3.2 Combination with Terrain Texture .............. 66
  6.3.3 N-Best Views ................................... 70
  6.3.4 Implementation Details .......................... 70
6.4 Conclusions ........................................ 73

6.1 Introduction

Digital elevation models are used ubiquitously within the geosciences, facilitating studies of natural and man-made phenomena across a wide range of scales. Commonly, elevation data, comprising height measurements linked by a grid or triangulation structure, are supplemented with digital image texture as the basis for qualitative and quantitative interpretation. Visualising and communicating terrain model data, with or without image texture, is important to fully exploit the benefits of geospatial data in geoscience applications. However, until now, user support for obtaining representative viewpoints and guiding the extraction of salient information about the terrain's shape has been minimal.

In this chapter, the information-theoretic framework for object understanding presented in Chapter 5 is applied to terrain visualization and terrain view selection. From a visibility channel between a set of viewpoints and the component polygons of a 3D terrain model, we obtain three specific polygonal information measures. These measures are used to visualize the information associated with each polygon of the terrain model. In order to enhance the perception of the terrain's shape, we explore the effect of combining the calculated information measures with the supplementary terrain texture. From polygonal information, we also introduce a method to select a set of representative views of the terrain model. Finally, we evaluate the performance of the
proposed techniques using example datasets. A publicly available framework for both
the visualization and the view selection of a terrain has been created.

The chapter is organized as follows. In Section 6.2, we review the concepts intro-
duced in Chapter 5. In Section 6.3, we apply three polygonal information measures to
3D terrain models and we present a combination of the polygonal measures with the
terrain texture and their corresponding viewpoint quality measures. We also present
a N best views method using the viewpoint quality measures presented in Chapter 5.
Finally, in Section 6.4, our conclusions are presented.

6.2 Background

As we have seen in Chapter 5, polygonal information measures $I_1$, $I_2$, and $I_3$ can be used
to quantify and visualize the information associated to a polygon of a 3D model. These
three measures provide us with different ways to perceive the model shape and they
are computed creating a visibility channel between a 3D model and a set of viewpoints.

In addition, in Chapter 5 we have presented several viewpoint quality measures
($VQ_1$, $VQ_2$, and $VQ_3$) using polygonal information measures $I_1$, $I_2$, and $I_3$. $VQ$s were
obtained by projecting (or spreading) the polygonal information of a 3D model to the
sphere of viewpoints. High values of $VQ$s correspond to viewpoints that see the most
complex parts of the model, which are represented by the areas with more occlusions
or significant details (i.e., with high values of polygonal information $I_1$ and $I_2$, and low
values of polygonal information $I_3$). On the other hand, low values of $VQ$s correspond
to viewpoints covering the smoothest areas of the model, with small changes in visibility
and less detail (i.e., with the lowest values of polygonal information $I_1$ and $I_2$, and the
highest values of polygonal information $I_3$).

6.3 Terrain Visualization

In this section, we show the results of applying the polygonal information measures and
the viewpoint quality measures presented in Chapter 5 to terrain visualization. First, the
values of polygonal measures $I_1$, $I_2$, and $I_3$ are visualized to enhance the perception of
a terrain model. Second, a method that combines the polygonal measures calculated
from the terrain model facets with additional image texture available in photorealistic
models is presented. Third, an $N$-best view selection method based on the viewpoint
quality measures $VQ_1$, $VQ_2$, and $VQ_3$ is shown. Finally, some implementation details
and the framework used in this chapter are presented.

6.3.1 Polygonal Information Visualization

As we have seen in Section 5.3.1, the polygonal information of a 3D model is created
from a visibility channel between the 3D model and a set of given viewpoints. Here we
compute the polygonal information measures of a terrain model.
To evaluate the applicability of the theoretical information measures on real data, a dataset is used. The dataset was collected for geological mapping purposes, and combines high resolution 3D geometry with digital imagery. While the 3D information provides an accurate geometric basis for quantitative analysis, integrating 2D images increases the geological information which may not only be present in the topographic description.

The dataset covers part of the Beckwith Plateau in the Book Cliffs, Utah, USA, and was acquired using a helicopter-based laser scanner (lidar) and co-mounted digital camera, as described by Buckley et al. [Buckley 2008] and Rittersbacher et al. [Rittersbacher 2014]. The laser data was collected using a Riegl Q240i-60 scanner and a Hasselblad HI 22mp camera was used for the image collection. With integrated dual frequency GNSS and inertial navigation systems, the dynamic platform position and orientations could be resolved to result in dense point cloud and digital image orientation data. The purpose of the data collection was to map sedimentary bodies, as part of geological outcrop studies [Rittersbacher 2014, Sima 2013a]. The 3D terrain model is formed by 834,212 triangles (see Figure 6.1).

To compute the polygonal information, the terrain model is centered in a sphere of 642 viewpoints built from the recursive discretization of an icosahedron, where the corresponding cameras are looking at the center of this sphere. To obtain the viewpoint sphere, the smallest bounding sphere of the model is computed and, then, the viewpoint sphere adopts the same center as the bounding sphere and a radius two times the radius of the model’s bounding sphere.

The polygonal information measures are computed from the projected area of every triangle within each viewpoint. These areas enable us to obtain the probability distributions of the visibility channel \(p(V), p(Z|V), \text{ and } p(Z)\) and, consequently the polygonal information measures \(I_1, I_2, \text{ and } I_3\) (Section 5.3.1).

In Figure 6.2, we show the polygonal information measures \(I_1\) (Figure 6.2 (row 1)), \(I_2\) (Figure 6.2 (row 2)), and \(I_3\) (Figure 6.2 (row 3)) corresponding to the terrain view shown in Figure 6.1 (a). To obtain these images, the polygonal information measures have been normalized between 0 and 1, and subtracted from 1. An exponential function
Chapter 6. Information Measures for Terrain Visualization

is applied to emphasize the polygonal information in the image. This function can be expressed as

\[ y = x^\alpha, \]  

(6.1)

where \( x \) is the original value, \( y \) is the value after applying the exponential function, and \( \alpha \) is a positive real value. Note that the values of polygonal information measures \( I_1, I_2, \) and \( I_3 \) are rendered with \( \alpha = 1 \) (original values), \( \alpha = 4, \) and \( \alpha = 8 \) (Figure 6.2, left to right). More contrasted images, resulting in sharper shading, are obtained with higher \( \alpha \) values.

Observe that low values of \( I_1 \) (Figure 6.2 (row 1)) and \( I_2 \) (Figure 6.2 (row 2)), corresponding to non-occluded triangles, are represented by bright colors (i.e., values near to 1) in the grey-map, while high values, corresponding to occluded triangles, are represented by dark colors (i.e., values near to 0) in the grey-map. On the other hand, low values of \( I_3 \), corresponding to triangles that are seen by viewpoints with low values of viewpoint information \( I_2 \), are represented by bright colors, and vice versa.

From information measure \( I_3 \) (Figure 6.2 (row 3)), we obtain a non-photorealistic visualization that permits us to perceive the shape of the terrain in a very detailed way, independent of lighting. Observe how the volume or the shape of the terrain is better perceived in this case than using measures \( I_1 \) or \( I_2 \).

### 6.3.2 Combination with Terrain Texture

As we have shown in the previous section, the shading based on \( I_1, I_2, \) and \( I_3 \) can help us to better understand the terrain model. In this section, the polygonal information is combined with the supplementary image texture to improve the perception of the terrain shape. From this combination, the perception of the terrain shape is enhanced without the need of the complex task of light positioning.

The terrain texture and the polygonal information are combined as

\[
C(p) = (1 - weight)k_d(p) + weight \ast k_d(p) \ast \frac{I'_i(z(p); V)}{l_n(\alpha)}
\]

(6.2)

to get the color of a pixel \( p \) of the screen. In this equation, \( weight \) is a value between 0 and 1 to weight, respectively, the combination of the original texture and the texture modified by the polygonal measure, \( k_d(p) \) is the color given by the terrain texture at the corresponding pixel \( p \), \( I'_i(z(p); V) \) is the scaled polygonal information at polygon \( z(p) \) (Section 5.4.1) where \( z(p) \) is the corresponding polygon at the pixel \( p \), and \( l_n(\alpha) \) is a light normalization factor to avoid the global darkening of the final image. To compute the exact value of the light normalization factor we sum the luminance of the image with original terrain texture. This value would change every time we move the viewpoint of the scene. To get an approximate value of \( l_n(\alpha) \) invariant to the viewpoint, \( l_n(\alpha) \) is computed as

\[
l_n(\alpha) = \frac{\sum_{z \in Z} d_z I'_i(z; V)^\alpha}{\sum_{z \in Z} d_z},
\]

(6.3)
Figure 6.2: Grey-map representation of polygonal information $I_1$ (row 1), $I_2$ (row 2), and $I_3$ (row 3) raised to 1 (column 1), 2 (column 2), and 4 (column 3). We have clipped the 0.1% of the data to do the normalization. From left to right, more contrasted images, resulting in sharper shading, are obtained with higher $\alpha$ values.
where $\bar{a}_z$ is the mean projected area of polygon $z$ from all the viewpoints. It is important to remark that in our framework this value can be easily adjusted by the user to obtain the desired luminance of the image. In Figure 6.3, we show the effect of combining the texture of the model with the polygonal information $I_1$, $I_2$, and $I_3$ powered to 8 and using $\text{weight} = 0.75$. In these images, the light normalization factor $l_n(\alpha)$ has been manually adjusted from the automatic value proposed by Equation 6.3 to obtain the final global luminance. In Figure 6.4, we show image enhancement done in image space using the GNU Image Manipulation Program (GIMP) compared to our enhancement. The GIMP enhancement has been done automatically using the Colors/Auto sub-menu with the options equalize (6.4j), white balance (6.4e), color enhance (6.4h), normalize (6.4g), stretch contrast (6.4i), and stretch HSV (6.4f). See how the polygonal information measures $I_1$, $I_2$, and $I_3$ emphasize better the shape of the terrain than the images obtained with GIMP. In addition note that the enhancement done with GIMP is computed in image space while the polygonal information measures enhancement is done in 3D space. This means that with the GIMP enhancement the coherence among different views can be lost. In Figure 6.5 we can see how a same part of the 3D model is painted with slightly different colors among different views.

Figure 6.3: Combination of a textured terrain with polygonal information $I_1$ (row 1), $I_2$ (row 2), and $I_3$ (row 3) powered to 4.
Figure 6.4: Image enhancement of the original image (a) in 3D space using polygonal measures (b, c, and d) and in image space using GIMP (e, f, g, h, i, and j).

Figure 6.5: Two different views of the same terrain model enhanced with GIMP with equalize option where some parts are painted with slightly different colors.
6.3.3 N-Best Views

In order to visualize a 3D terrain model, we are interested in selecting a set of representative views which provides an adequate understanding of the terrain. With this goal in mind, we present a new viewpoint selection algorithm inspired by the viewpoint selection algorithms presented in Sima et al. [Sima 2013b] and in the Chapter 5.

Our N-best views algorithm is based on a two-step iterative process that computes a successive set of views by taking only into account at each iteration the previously unseen triangles. First, the algorithm computes the best view in terms of the viewpoint quality measures presented in Section 5.4.1. Second, the algorithm discards the triangles with projected area at least 50% of the maximum projected area captured from all the viewpoints. Then, the best view is again selected without taking into account the discarded triangles. The process is repeated until the algorithm achieves a predefined number of representative views or a percentage of terrain coverage. The quality measures used by this algorithm are, respectively, $\text{VQ}_1$, $\text{VQ}_2$, or $\text{VQ}_3$ (Equation 5.18).

In Figure 6.6, we show the results of selecting six representative views using $\text{VQ}_1$ (row 1), $\text{VQ}_2$ (row 2), and $\text{VQ}_3$ (row 3) from the sphere of viewpoints used to compute the polygonal information measures $I_1$, $I_2$, and $I_3$. Under each image we show the percentage of the terrain covered with the first N-best views. The first view (column 1) is the one that provides the maximum information about the model according to Equation 5.18. Successively, each new image adds new information to the previous images. Observe the similarity of the images obtained from the three different viewpoint quality measures. Note that, in this example, after the fifth view, few new triangles are added and also little information (see Figure 6.8). In Figure 6.7 we see the results obtained from the set of viewpoints provided by the user and that have been used to obtain the 3D model. These viewpoints correspond to the ones that have been used to take the photos of the terrain in order to create the final terrain model texture. Note that, in this situation, the views are less global than in the previous case (Figure 6.6) and, thus, the triangles are discarded more slowly and it takes more images before they are repeated.

6.3.4 Implementation Details

All the tests carried out in this chapter have been done in a framework that is publicly available at https://github.com/limdor/quoniam-terrain. This framework allows the user to visualize a textured 3D terrain model in three different ways: textured with the original terrain texture, only the polygonal information of the terrain, or a combination of both. In the last two options the user can change the $\alpha$-value of the exponential function. The user can also change both the weight and the light normalization factors for the combination of the original textured terrain with the polygonal information measures.

Our tests were run on a Intel® Core™ i7-2600K 3.40GHz machine with 16 GB RAM and an ATI Radeon™ HD 6950 with 2048 MB. The time of the preprocessing step, that is, the cost to compute the projected areas $a_z(v)$ and $a_t(v)$ for the Book Cliffs terrain
6.3. Terrain Visualization

Figure 6.6: Six best views of the Book Cliffs terrain using $VQ_1$ (row 1), $VQ_2$ (row 2), and $VQ_3$ (row 3) and discarding triangles seen with a quality better than 50%. These views have been obtained from the sphere of viewpoints.

Figure 6.7: Six best views of the Book Cliffs terrain using $VQ_1$ (row 1), $VQ_2$ (row 2), and $VQ_3$ (row 3) and discarding triangles seen with a quality better than 50%. These views have been obtained from the set of views provided by the helicopter to create the 3D model.
Figure 6.8: Information gained with each new viewpoint selected for $VQ_1$ (row 1), $VQ_2$ (row 2), and $VQ_3$ (row 3) using the sphere of viewpoints (column 1) and the set of views provided by the helicopter to texture the 3D model (column 2).
model with 834212 triangles took 47959 ms. In order to get a good precision, we have used a projection resolution of 4096 × 4096.

6.4 Conclusions

In this chapter, we have applied the information theory framework for object understanding presented in Chapter 5 to terrain visualization. Defining a visibility channel between a set of viewpoints and the polygons of a terrain model, we have obtained and visualized the polygonal information getting several shading approaches. We have also presented a combination of the polygonal information measures and the original terrain texture in order to enhance the perception of the terrain shape. Finally, we have used the viewpoint quality measures $VQ_1$, $VQ_2$, and $VQ_3$ presented in Chapter 5 to get a set of representative views of the terrain.
CHAPTER 7

3D Shape Retrieval Using Viewpoint Information-Theoretic Measures

Contents

7.1 Introduction ......................................................... 75
7.2 Background .......................................................... 76
  7.2.1 Feature-Based Methods ....................................... 77
  7.2.2 Graph-Based Methods ......................................... 77
  7.2.3 View-Based Methods .......................................... 77
7.3 View-Based Similarity Framework ................................. 78
  7.3.1 $L_2$ Distance between Information Spheres .............. 78
  7.3.2 Earth Mover's Distance between Information Histograms ... 80
  7.3.3 Mutual Information Difference .............................. 81
7.4 Results and Discussion ............................................. 81
  7.4.1 Experimental Results ......................................... 81
  7.4.2 Discussion ..................................................... 86
7.5 Conclusions ......................................................... 88

7.1 Introduction

Quantifying the shape similarity between 3D polygonal models is a key problem in different fields, such as computer graphics, computer vision and pattern recognition. Recently, as the number of large digital repositories of 3D models grows dramatically, 3D data are becoming ubiquitous. As a result, there is an increasing demand for search engines that are able to retrieve similar models using shape similarity measures. In the last few years, a number of algorithms have been proposed for the retrieval of both rigid (see [Tangelder 2008] for a survey on content-based 3D shape retrieval) and non-rigid 3D shapes [Lian 2013]. Several 3D shape-based retrieval methods are based on view similarity, where two 3D models are considered similar if they look similar from all viewing angles. In this chapter, we advance in this line by tackling the shape similarity problem from an information-theoretic framework.
From this framework, several information-theoretic methods are presented to compute the similarity matrix of a set of models. Given a 3D model, our information measures are obtained from a visibility channel created between the set of viewpoints and the polygonal mesh. We define different information measures: the mutual information of a 3D model, and the specific information measures $I_1$ and $I_2$ associated with each viewpoint. The last-mentioned measures correspond to two different forms of decomposing the mutual information and enable us to create two different information spheres for each model.

From the above information-theoretic measures, we present three methods to obtain the similarity matrix for all the models of a database. In the first method, a registration process between the information spheres of two models is carried out to obtain the pose that achieves the minimum $L_2$ distance. This distance quantifies the degree of dissimilarity between two model shapes. In the second method, the earth mover's distance between the information histograms of two models is also used to calculate the degree of dissimilarity between the corresponding shapes. In the third method, the mutual information of a 3D model is considered as a shape signature and the difference in absolute value of the mutual information of each model can be also seen as a shape discrimination measure.

This chapter is organized as follows. In Section 7.2, we summarize some related work in 3D shape retrieval. In Section 7.3, we propose several methods to compute the similarity between 3D models. In Section 7.4, experimental results show the performance of the proposed measures for 3D shape retrieval. Finally, in Section 7.5, the conclusions are presented.

### 7.2 Background

Recent developments in techniques for modeling, digitizing and visualizing 3D shapes have provoked an explosion in the number of available 3D models on the Internet and in specific databases. This has led to the development of 3D shape retrieval systems (see [Tangelder 2008] for a survey) that, given a query object, retrieve similar 3D objects.

At conceptual level, a typical shape retrieval framework consists of a database with an index structure created off-line and an on-line query engine. Each 3D model has to be identified with a shape descriptor, providing an overall description of its shape. The indexing data structure and the searching algorithm are used to carry out an efficient search. The on-line query engine computes the query descriptor, and the models similar to the query model are retrieved by matching descriptors to the query descriptor from the index structure of the database. The similarity between two descriptors is quantified by a dissimilarity measure.

According to Tangelder and Veltkamp [Tangelder 2008], 3D shape retrieval systems are usually evaluated with respect to several requirements of content based 3D retrieval, such as shape representations requirements, properties of dissimilarity measures, efficiency, discrimination abilities, ability to perform partial matching, robustness, and
7.2. Background

necessity of pose normalization. Different number of tools exist to validate 3D shape retrieval systems such as the Princeton Shape Benchmark (PSB) [Shilane 2004], the Purdue engineering shape benchmark [Jayanti 2006], or the McGill 3D shape benchmark [Siddiqi 2008]. There is also the SShape REtrieval Contest (SHREC) organized every year since 2006 [Veltkamp 2006] where a dataset is provided to the participants to run their 3D shape retrieval methods. Some of the above benchmarks have been part of SHREC editions.

Shape matching methods can be divided in three broad categories: feature-based methods, graph-based methods, and view-based methods.

7.2.1 Feature-Based Methods

Feature-based methods are the most commonly used as features that can directly denote the geometric and topological properties of 3D shapes. According to the type of shape features used, feature based methods can be further categorized into: global features, global feature distributions, spatial maps and local features, in all of which two models are compared according to their feature distance in the fixed d-dimensional space. The first three categories use a single d-dimensional vector to represent features, while local feature-based methods compute feature vectors for a number of surface points, which are often the salient points of a 3D model. Corresponding feature based methods of each category for the 3D shape retrieval include self-similarity (symmetry) [Kazhdan 2004] and the global descriptors based on volume and area [Zhang 2001], distance distributions [Osada 2002] and spectral shape analysis [Reuter 2006, Lian 2013], statistical moments [Kazhdan 2003, Novotni 2003], and the local features combined with the bag-of-words model [Bronstein 2011] or the heat kernel diffusion [Sun 2009].

7.2.2 Graph-Based Methods

Graph-based methods use a graph to extract a geometric meaning from a 3D shape and utilize the topological information of 3D objects to measure the similarity between them, rather than only considering the pure geometry of the shape as the feature-based methods do. Graph-based methods can be applied to articulated models. According to the type of the used graphs, three categories can be considered in this technique including model graphs [El-Mehalawi 2003b, El-Mehalawi 2003a], Reeb graphs [Hilaga 2001, Tung 2005], and skeletons [Sundar 2003]. Compared with the feature based methods, the graph based methods are less robust, but the graph based structure is suitable for partial matching.

7.2.3 View-Based Methods

Based on the fact that 3D models are similar when they look similar from all viewing angles, view-based similarity methods are proposed. The earlier reference to view-based retrieval is given by [Loffler 2000] who used a 2D query interface to retrieve 3D models. Funkhouser et al. describe an image-based approach allowing users to query the
engine by drawing one or more sketches [Funkhouser 2003]. Chen et al. provide a system-based on a set of lightfield descriptors [Chen 2003], and one hundred orthogonal projections of an object are encoded both by Zernike moments and Fourier descriptors as features for retrieval. Gonzalez et al. use the sphere of viewpoints with viewpoint mutual information as a descriptor of the model [González 2007]. The sketch-based 3D model retrieval system proposed by [Yoon 2010] is robust against variations of shape, pose or partial occlusion of the sketches, but the drawing process is still a little botherening. Although the discriminative and robust sketch-based 3D shape retrieval system by [Shao 2011] requires dense sampling and registration and incurs a high computational cost, critical acceleration methods based on pre-computation and multi-core platforms or GPUs are designed to achieve interactive performance. Eitz et al. collect a significant number of sketches for the evaluation of shape retrieval performance and achieve significantly better result than the previous methods [Eitz 2012], but realistic inputs are still a very hard problem, which is related to their bag-of-features and the new descriptor for line-art renderings. Liu et al. design a statistical measure based on sketch similarity for CAD model retrieval, which accounts for users' drawing habits [Liu 2013]. The limitation of the method is that a single freeform sketch mainly captures some geometric information other than semantic meanings.

7.3 View-Based Similarity Framework

As we have seen in Sections 2.4 and 3.3, \( I(V; Z) \) expresses the degree of correlation between a set of viewpoints and the 3D model, and the viewpoint information measures \( I_1(v; Z) \) and \( I_2(v; Z) \) quantify the degree of correlation between a single viewpoint and the model. Our view-based similarity approach is an extension of the method presented by [González 2007], where preliminary results were given for a single measure, viewpoint mutual information.

In this section, we present three different methods to evaluate the shape similarity between two models and to obtain the distance matrix for all the models of a data set. These methods are respectively based on the \( L_2 \) distance between information spheres (Section 7.3.1), the earth mover’s distance between information histograms (Section 7.3.2), and the absolute difference between the mutual information of each model (Section 7.3.3).

7.3.1 \( L_2 \) Distance between Information Spheres

In this method, given a 3D model, two information spheres are respectively obtained by computing the information measures \( I_1 \) and \( I_2 \) for each viewpoint presented in Section 3.3. These measures have been defined as follows:

- The viewpoint information \( I_1 \) of a viewpoint \( v \) is defined as

\[
I_1(v; Z) = \sum_{z \in Z} p(z|v) \log \frac{p(z|v)}{p(z)}. \tag{7.1}
\]
7.3. View-Based Similarity Framework

Figure 7.1: (first row) $I_1$-spheres and (second row) $I_2$-spheres corresponding to four different 3D models of the same class.

- The viewpoint information $I_2$ of a viewpoint $v$ is defined as

$$
I_2(v; Z) = H(Z) - H(Z|v) = -\sum_{z \in Z} p(z) \log p(z) + \sum_{z \in Z} p(z|v) \log p(z|v).
$$

Then, a registration process is done to find the minimum distance that characterizes the degree of dissimilarity between two models.

Figure 7.1 shows both the $I_1$-spheres and the $I_2$-spheres for four different models of the same class. Observe how similar information spheres are obtained for all the models, although the most similar patterns are provided by the measure $I_1$. The information spheres are considered as shape descriptors (or signatures) for a given model.

To compute the dissimilarity (or distance) between two $I_1$-spheres (or $I_2$-spheres), a registration process is carried out to obtain the pose that achieves the minimum distance between the viewpoint information values. In the registration process, we aim to find the transformation that brings one sphere (floating) into the best possible spatial correspondence with the other one (fixed) by minimizing the distance between the information measures of the corresponding viewpoints. The distance used is the $L_2$ distance which is based on the absolute difference between each pair of matching viewpoint information values.

To achieve the best matching between both the fixed and the floating sphere, we consider the following points. First, the discrete nature of our information spheres (e.g., 642 viewpoints) requires an interpolator component. In our implementation, the nearest neighbor interpolator has been used. Second, the $L_2$ distance between the information spheres $S_1$ and $S_2$ (corresponding to the models $Z_1$ and $Z_2$) for a specific matching is given by

$$
D(S_1, S_2) = \sqrt{\sum_{v \in \mathcal{V}} (I(v; Z_1) - I(v; Z_2))^2},
$$

where $I(v; Z)$ stands for $I_1(v; Z)$ or $I_2(v; Z)$. Third, we use two transformation parameters (degrees of freedom): $R(\theta)$ and $R(\varphi)$, defined respectively as the rotation around $Z$ and $Y$ axis. These two parameters take values in the range $[0^\circ, 360^\circ]$ and $[0^\circ, 180^\circ]$. 
respectively. Through this process we get the method to be robust to rotations of the models.

In this method, we assume that the correct matching is given by the minimum value of $D(S_1, S_2)$. Since this matching process is time-consuming if all the possible positions are checked, we use Powell’s method to speed up the registration [Powell 1964]. Powell’s method is a numerical optimizer that finds the minimum of a function without using derivatives.

To sum up, the fundamental idea of this view-based similarity approach is that the viewpoint measures used to build the information spheres supply an information measure for each viewpoint. Thus, the sphere of viewpoints can be seen as a shape representation of the object. In our case, two 3D models are similar when their corresponding information spheres are also similar, that is, capture a similar information distribution. Note that we only store one scalar value for each viewpoint, differently to other methods, that store the silhouette or the depth map [Eitz 2012, Ohbuchi 2008].

### 7.3.2 Earth Mover’s Distance between Information Histograms

As in Section 7.3.1, the first step is the creation of the $I_1$ and $I_2$ spheres corresponding to a given model. Then, we obtain the information histograms that are used as shape descriptors of the model.

To create both the $I_1$-histogram and the $I_2$-histogram from the corresponding information spheres, we need to fix three parameters: the minimum and the maximum value of the information measure (i.e., $I_1$ or $I_2$), and the number of bins of the histogram. Taking into account that $I_1$ is always greater or equal to 0, its minimum value has been fixed to 0. On the other hand, the maximum value of $I_1$ has been taken from the highest $I_1$ value among all the models in the database. In a similar way, the minimum and maximum values of $I_2$ have been obtained from the lowest and highest $I_2$ values among all the models in the database, respectively. The maximum and minimum values could be also fixed using a training set and doing some kind of clipping to avoid outliers. As we will see in the next section, our tests have been performed using different number of bins.

The dissimilarity between two models is computed by the Earth Mover’s Distance (EMD) between their histograms [Rubner 1998]. EMD is a measure of the distance between two distributions. If the two distributions are interpreted like two ways of piling up an amount of earth, then EMD is the least amount of work needed to turn one pile into the other. A unit of work corresponds to transporting a unit of earth by a unit of distance. In our case, this distance is given by the distance between bins and the amount of earth is given by the probability of belonging to a bin. If both distributions have the same amount of earth, EMD is a true distance. This condition is also fulfilled in our case.

The earth mover’s distance between two information histograms $H_1$ and $H_2$ is defined as

$$
EMD(H_1, H_2) = \frac{\sum_{i \in H_1} \sum_{j \in H_2} c_{ij} f_{ij}}{\sum_{i \in H_1} \sum_{j \in H_2} f_{ij}},
$$

(7.4)

7.4. Results and Discussion

In this section, we analyze and discuss the behavior of $I_1$ and $I_2$ spheres, $I_1$ and $I_2$ histograms, and $I(V; Z)$ as shape descriptors for 3D object retrieval.

7.4.1 Experimental Results

To calculate the information-theoretic measures presented in Section 3.3, we need to obtain the projected area of every polygon for every viewpoint, and these areas will enable us to obtain the probabilities of the visibility channel ($p(V)$, $p(Z|V)$, and $p(Z)$). In this chapter, all the measures have been computed without taking into account the background, and using a projection resolution of $640 \times 640$. In our experiments, all the models are centered inside a sphere of 642 viewpoints built from the recursive discretization of an icosahedron and the camera is looking at the center of this sphere. To obtain the viewpoint sphere, the smallest bounding sphere of the model is obtained and, then, the viewpoint sphere adopts the same center as the bounding sphere and a radius three times the radius of the bounding sphere. Centering the object to the center of the sphere we get a method invariant to translations and, as the viewpoints are uniformly distributed over the sphere, we have also invariance to rotations. The 642 values of the viewpoint sphere for $I_1$ and $I_2$, and the mutual information of the

where $c_{ij}$ represents the distance between bin $i$ of histogram $H_1$ and bin $j$ of histogram $H_2$, and $f_{ij}$ represents the amount of occurrences that is transferred between bin $i$ and bin $j$.

7.3.3 Mutual Information Difference

We can also use the mutual information $I(V; Z)$ between the set of viewpoints and the model as a signature of the model. Let us remember that the mutual information expresses the degree of correlation or dependence between the set of viewpoints and the model. The distance between two models is now computed as the difference between their mutual information in absolute value. This is a very coarse approach but the advantage is that the signature is represented by a single scalar value and the comparison between signatures is really fast. This would allow, at almost no cost, to build a short list of candidate matching models. In Figure 7.2, we see one example that shows how the objects of a same class have similar values of $I(V; Z)$.

Figure 7.2: 3D models of the same class with similar value of $I(V; Z)$. 

![Figure 7.2: 3D models of the same class with similar value of $I(V; Z)$.
](image-url)
Table 7.1: For each measure, the size of the signature, the time to generate it, and the time to compare two models are shown.

<table>
<thead>
<tr>
<th>Descriptors</th>
<th>Size (elements)</th>
<th>Generation time (s)</th>
<th>Comparison time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1, I_2$-sphere</td>
<td>642</td>
<td>1.44</td>
<td>0.064236</td>
</tr>
<tr>
<td>$I_1, I_2$-histogram</td>
<td># of bins</td>
<td>1.44</td>
<td>0.001407</td>
</tr>
<tr>
<td>$I(V; Z)$</td>
<td>1</td>
<td>1.36</td>
<td>0.000000</td>
</tr>
</tbody>
</table>

visibility channel are used to compute the shape descriptors of every model as explained in Section 7.3.

The shape descriptors of each model are computed in advance and stored into a database. At run time, we only have to compare the shape descriptors of the models and to generate a descriptor when a new model is added to the database. The storage size for each descriptor in number of float values depends on the method used as well as the cost of computing each descriptor and the comparison between descriptors (see Table 7.1). Our tests were run on an Intel® Core™ i7-2600K 3.40GHz machine with 16 GB RAM and an ATI Radeon™ HD 6950 with 2048 MB.

To test the performance of our methods we use the Princeton Shape Benchmark (PSB) database and its utilities [Shilane 2004]. First, we run the methods using the training set of 907 objects using the base classification file that groups the models in 90 different classes. The training set is intended to tune the parameters of the methods, in our case we only have to adjust the number of bins when we create the histograms. In Table 7.2 we can see the results of executing our three methods with this training data set. For the method that uses the information histograms we have tested with different number of bins: 16, 32, 64, 96 and 128.

The first three statistics (nearest neighbor (NN), first tier (FT), and second tier (ST)) indicate the percentage of the top $K$ matches that belong to the same class as the query. For the nearest neighbor statistic, $K$ is 1, and for the first tier and second tier statistics, $K$ is $C - 1$ and $2(C - 1)$, respectively, where $C$ is the size of the query’s class. In all three cases, an ideal matching result (where all the other models within the query’s class appear as the top matches) gives a score of 100%. The fourth statistic is the E-Measure (E-M), which is a composite measure of precision and recall for a fixed number of retrieved results. The E-Measure is defined by

$$E = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

where $P$ is the precision and $R$ is the recall. Remember that $P = TP/(TP + FP)$ and $R = TP/(TP + FN)$ where $TP$ are the true positives, $FP$ the false positives, and $FN$ the false negatives. The maximum value of $E$ is 1 and higher values indicate better results.

The fifth statistic (discounted cumulative gain) (DCG) gives a sense of how well the overall retrieval would be viewed by a human. Correct shapes near the front of the list are more likely to be seen than correct shapes near the end of the list. More information
Results and Discussion

<table>
<thead>
<tr>
<th>Measures</th>
<th>NN</th>
<th>FT</th>
<th>ST</th>
<th>E-M</th>
<th>DCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$-sphere</td>
<td>47.5%</td>
<td>24.7%</td>
<td>32.8%</td>
<td>15.5%</td>
<td>49.8%</td>
</tr>
<tr>
<td>$I_2$-sphere</td>
<td>29.7%</td>
<td>13.8%</td>
<td>19.1%</td>
<td>9.7%</td>
<td>37.9%</td>
</tr>
<tr>
<td>$I_1$-hist 96</td>
<td>25.5%</td>
<td>12.2%</td>
<td>18.1%</td>
<td>10.3%</td>
<td>37.3%</td>
</tr>
<tr>
<td>$I_1$-hist 128</td>
<td>24.8%</td>
<td>12.0%</td>
<td>18.1%</td>
<td>10.3%</td>
<td>37.2%</td>
</tr>
<tr>
<td>$I_1$-hist 64</td>
<td>24.6%</td>
<td>12.0%</td>
<td>17.8%</td>
<td>10.2%</td>
<td>37.1%</td>
</tr>
<tr>
<td>$I_1$-hist 32</td>
<td>23.0%</td>
<td>12.0%</td>
<td>17.6%</td>
<td>10.1%</td>
<td>36.6%</td>
</tr>
<tr>
<td>$I_1$-hist 16</td>
<td>18.6%</td>
<td>10.7%</td>
<td>16.5%</td>
<td>9.6%</td>
<td>35.2%</td>
</tr>
<tr>
<td>$I_2$-hist 128</td>
<td>17.4%</td>
<td>9.0%</td>
<td>14.2%</td>
<td>8.1%</td>
<td>33.1%</td>
</tr>
<tr>
<td>$I_2$-hist 96</td>
<td>17.0%</td>
<td>9.2%</td>
<td>14.1%</td>
<td>8.1%</td>
<td>33.2%</td>
</tr>
<tr>
<td>$I_2$-hist 64</td>
<td>16.9%</td>
<td>8.8%</td>
<td>14.5%</td>
<td>8.0%</td>
<td>33.1%</td>
</tr>
<tr>
<td>$I_2$-hist 32</td>
<td>16.9%</td>
<td>8.7%</td>
<td>13.3%</td>
<td>8.0%</td>
<td>32.7%</td>
</tr>
<tr>
<td>$I_2$-hist 16</td>
<td>14.3%</td>
<td>7.9%</td>
<td>12.0%</td>
<td>7.4%</td>
<td>31.4%</td>
</tr>
<tr>
<td>$(V;Z)$</td>
<td>6.8%</td>
<td>4.5%</td>
<td>8.1%</td>
<td>4.7%</td>
<td>27.6%</td>
</tr>
</tbody>
</table>

Table 7.2: Results of our measures with the Princeton Shape Benchmark training set.

<table>
<thead>
<tr>
<th>Measures</th>
<th>NN</th>
<th>FT</th>
<th>ST</th>
<th>E-M</th>
<th>DCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I_1$-sphere</td>
<td>39.4%</td>
<td>20.8%</td>
<td>27.9%</td>
<td>14.4%</td>
<td>45.3%</td>
</tr>
<tr>
<td>$I_2$-sphere</td>
<td>27.6%</td>
<td>12.5%</td>
<td>17.6%</td>
<td>9.3%</td>
<td>36.3%</td>
</tr>
<tr>
<td>$I_1$-hist 96</td>
<td>18.2%</td>
<td>8.9%</td>
<td>14.0%</td>
<td>8.5%</td>
<td>32.9%</td>
</tr>
<tr>
<td>$I_2$-hist 96</td>
<td>14.0%</td>
<td>6.9%</td>
<td>11.4%</td>
<td>6.8%</td>
<td>30.2%</td>
</tr>
<tr>
<td>$(V;Z)$</td>
<td>4.0%</td>
<td>2.6%</td>
<td>5.0%</td>
<td>3.6%</td>
<td>25.0%</td>
</tr>
</tbody>
</table>

Table 7.3: Results of our measures with the Princeton Shape Benchmark test set.

about these statistics can be seen at [Shilane 2004].

In Table 7.2, the methods have been ordered using the NN statistic. Observe that the best results are obtained with the $L_2$ distance between $I_1$-spheres, which are clearly better than the ones obtained between $I_2$-spheres. Thus, these results confirm the visual hint (see Figure 7.1) that $I_1$-spheres are better descriptors than $I_2$-spheres. Concerning the information histograms, the EMD distance also achieves better results with $I_1$ than with $I_2$. We can also see that the best results with the information histograms are obtained using 96 bins. The $I_2$-histogram method with 128 bins gives slightly better results with the nearest neighbor and the second tier statistic but not with the first tier. Thus, from now on, we will use 96 bins for the histogram-based methods. Finally, for illustrative purposes, we have also added the mutual information difference, which being a scalar measure has a very low discrimination power.

Once we have fixed the number of bins, we analyze the behavior of our approach with the PSB test database that contains 907 objects distributed in 92 different classes. In Table 7.3, we can observe how the order of the measures is kept with respect to the training data set although the results have worsened. Next, we analyze these results.
Chapter 7. 3D Shape Retrieval Using Viewpoint Measures

Figure 7.3: Two objects with similar $I(V; Z)$ values (1.58079 and 1.58275) and different patterns for the $I_1$-spheres.

Figure 7.4: 3D models of the same class where we can see some models where the value of $I(V; Z)$ is quite different between them. The three last models are malformed.

The measure with clearly worst results is $I(V; Z)$ but if we go deeper we can observe why it fails and when this measure could be useful. As we can see in Figure 7.2, the objects of the same class tend to have a similar value of $I(V; Z)$. However, objects of different classes can also have a similar value of $I(V; Z)$. When in this case we check the information spheres, we can see that their patterns are considerably different (see Figure 7.3). That is, even though the distribution of the $I_1$ values on the sphere can be different, their average can be similar. Taking into account this behavior, the $I(V; Z)$ method could be used as a filter to select a subset of models with similar $I$ value and, then, other methods such the ones based on information histograms or information spheres could be applied.

In some occasions we can see a class with the values of $I(V; Z)$ not so similar as expected (see Figure 7.4). To explain this behavior we analyze some models with an unexpected value of $I(V; Z)$. In Figure 7.5 we can see a model where $I(V; Z)$ is different from other models of the same class. The model has been rendered in a way that the background is white, the faces seen from the front are gray, and the faces seen from the rear are black. For many purposes, when you can see the back face of polygons, it is considered that the model is malformed. This is due to the fact that if we apply the
7.4. Results and Discussion

Figure 7.5: Malformed model where we can see the back face of some polygons in black and the background through some holes in white. This model corresponds to the second model of Figure 7.4.

Figure 7.6: (left) $I_1$-histogram and (right) $I_2$-histogram corresponding to two models of the same class (see Figure 7.8).

back face culling optimization then the polygon seen from the rear are invisible. If we do not apply the back face culling, we can have problems with the normals when we apply illumination methods.

These malformed models affect the performance of our measure due to the way used to compute the visibility channel. To compute the projected area of a polygon from a viewpoint, we project the polygons from the front and from the back. This is done to handle when a model has wrong normals and mixed polygons clockwise and counter clockwise as it happens in PSB. This implies that if we have a plane constructed with only two triangles instead of four (two looking up and two looking down), we construct a channel where the two triangles are seen by all the viewpoints. To get a good behavior of the measures we need that the half hemisphere of viewpoints see two triangles looking up and the other half see two other triangles looking down.

Figures 7.6 and 7.7 show the $I_1$ and $I_2$ histograms of two models of the same class and of two models of different classes, respectively. Observe that we have reduced the range of the bins to enhance the visualization. In Figure 7.8, we can see the models used to create the histograms. Basically, we can observe how the histograms of two models of the same class (Figure 7.6) have a remarkable similarity, while the histograms of two models of very different classes (Figure 7.7) are very dissimilar.

In Table 7.3 we can also observe that the performance of measures has decreased with relation to the ones for the training set shown in Table 7.2 but the order of efficiency is preserved between the methods. This behavior can be explained if we analyze the results in the training and test set grouped by classes. If we take the two classes with the worst results in the test set, we see that they are the covered wagons and the
Figure 7.7: (left) $I_1$-histogram and (right) $I_2$-histogram corresponding to two models of different classes (see Figure 7.8).

Figure 7.8: 3D models used for the histogram figures.

 variants for satellite dish classes, two classes that are not present in the training set. It also happens that the class that gives better results in the training set is the swings set which is not present in the test set. If we look at the classes that are common in both training and test sets we can see that sometimes the result is better with the training set and sometimes is better with the test set as we would expect.

A possible explanation of the bad results for the covered wagons class is that more than half of the models are malformed. For the satellite dish class, the overall shape of the models is different and our methods are not able to detect the common piece that is the dish (see Figure 7.9). It also happens that the dish of model m1813 is malformed.

7.4.2 Discussion

The results for our 3D shape retrieval framework based on viewpoint information channel are preliminary, and from Table 7.4 we can see that even with our best method, based on $I_1$-spheres, they are still far from being competitive. They depend heavily on the good construction of the models and, hence, we plan to check with databases of

Figure 7.9: The satellite dish class of the test set.
<table>
<thead>
<tr>
<th>Shape Descriptor</th>
<th>Storage Size (bytes)</th>
<th>Generation Time (s)</th>
<th>Comparison Time (s)</th>
<th>Nearest Neighbor (%)</th>
<th>First Tier (%)</th>
<th>Second Tier (%)</th>
<th>E-Measure (%)</th>
<th>DCG (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFD</td>
<td>4,700</td>
<td>3.25</td>
<td>0.001300</td>
<td>65.7</td>
<td>38.0</td>
<td>48.7</td>
<td>28.0</td>
<td>64.3</td>
</tr>
<tr>
<td>REXT</td>
<td>17,416</td>
<td>2.22</td>
<td>0.000229</td>
<td>60.2</td>
<td>32.7</td>
<td>43.2</td>
<td>25.4</td>
<td>60.1</td>
</tr>
<tr>
<td>SHD</td>
<td>2,184</td>
<td>1.69</td>
<td>0.000027</td>
<td>55.6</td>
<td>30.9</td>
<td>41.1</td>
<td>24.1</td>
<td>58.4</td>
</tr>
<tr>
<td>GEDT</td>
<td>32,776</td>
<td>1.69</td>
<td>0.000450</td>
<td>60.3</td>
<td>31.3</td>
<td>40.7</td>
<td>23.7</td>
<td>58.4</td>
</tr>
<tr>
<td>EXT</td>
<td>552</td>
<td>1.17</td>
<td>0.000008</td>
<td>54.9</td>
<td>28.6</td>
<td>37.9</td>
<td>21.9</td>
<td>56.2</td>
</tr>
<tr>
<td>SECSHEL</td>
<td>32,776</td>
<td>1.38</td>
<td>0.000451</td>
<td>54.6</td>
<td>26.7</td>
<td>35.0</td>
<td>20.9</td>
<td>54.5</td>
</tr>
<tr>
<td>VOXEL</td>
<td>32,776</td>
<td>1.34</td>
<td>0.000450</td>
<td>54.0</td>
<td>26.7</td>
<td>35.3</td>
<td>20.7</td>
<td>54.3</td>
</tr>
<tr>
<td>SECTORS</td>
<td>552</td>
<td>0.90</td>
<td>0.000014</td>
<td>50.4</td>
<td>24.9</td>
<td>33.4</td>
<td>19.8</td>
<td>52.9</td>
</tr>
<tr>
<td>CEGI</td>
<td>2,056</td>
<td>0.37</td>
<td>0.000027</td>
<td>42.0</td>
<td>21.1</td>
<td>28.7</td>
<td>17.0</td>
<td>47.9</td>
</tr>
<tr>
<td>$l_1$-sphere</td>
<td>2,568</td>
<td>1.44</td>
<td>0.064236</td>
<td>39.4</td>
<td>20.8</td>
<td>27.9</td>
<td>14.4</td>
<td>45.3</td>
</tr>
<tr>
<td>EGI</td>
<td>1,032</td>
<td>0.41</td>
<td>0.000014</td>
<td>37.7</td>
<td>19.7</td>
<td>27.7</td>
<td>16.5</td>
<td>47.2</td>
</tr>
<tr>
<td>D2</td>
<td>136</td>
<td>1.12</td>
<td>0.000002</td>
<td>31.1</td>
<td>15.8</td>
<td>23.5</td>
<td>13.9</td>
<td>43.4</td>
</tr>
<tr>
<td>SHELLS</td>
<td>136</td>
<td>0.66</td>
<td>0.000002</td>
<td>22.7</td>
<td>11.1</td>
<td>17.3</td>
<td>10.2</td>
<td>38.6</td>
</tr>
</tbody>
</table>

Table 7.4: Results from [Shilane 2004], where we have merged the results of our $l_1$-sphere method (see Table 7.3).
well formed models. But we should also investigate a strategy to overcome this problem, maybe by projecting the triangles as a double face. We are also limited in principle to rigid models, but as far as they can identify the different poses of a same model as coming from the same family the retrieval would be correct.

A multilevel retrieval strategy could be used within our framework. We will first use MI to build at practically no cost a preliminary list of candidate models, which would be further refined into a shortlist with histogram comparison at very low cost. Finally, we will register the short list elements with the \( I_1 \)-spheres.

We have used the standard \( L_2 \) distance as a registration measure for 3D spheres, but maybe a more conceptual measure would yield better results. An experiment with humans classifying objects by only looking at the information spheres would give the maximum discrimination power of our measures, and thus the room for improvement in a registration measure.

### 7.5 Conclusions

In this chapter, we have presented a framework for 3D shape retrieval based on the information channel between the set of viewpoints around a 3D model and the 3D model polygons. From this channel we have derived different similarity measures, based on the decomposition of mutual information. The presented quality measures associated with the sphere of viewpoints have been used as a shape representation of the object. The performance of these measures has been tested using the Princeton Shape Benchmark database obtaining the best results with the registration of \( I_1 \)-spheres using \( L_2 \) distance. Used individually, our measures can not compete with the state of the art methods, but offer room for a multilevel retrieval strategy where mutual information would be used to obtain a preliminary list of candidates.
8.1 Contributions

The main objective of this thesis was to find good information-theoretic measures to improve the perception of 3D polygonal models and their recognition. This objective has been achieved with the following contributions:

- We have analyzed the use of several mutual information decompositions of an information channel between a set of viewpoints and a 3D model to quantify the quality of a viewpoint. Two measures of specific information introduced in the field of neural systems have been applied to quantify the information associated with a viewpoint. These measures have been compared with viewpoint entropy and viewpoint mutual information, and different experiments have shown their performance in best view selection. The concepts of surprise, diversity, and informativeness associated with a viewpoint have been also discussed.

  This contribution has been published in Proceedings of 21st GraphiCon International Conference on Computer Graphics and Vision, pages 16–19, September 2011 titled Viewpoint Information. [Bonaventura 2011]

- We have analyzed the performance of the most significant viewpoint quality measures presented in the literature and we have grouped all of them together in a common framework.

  We have reviewed the main measures for viewpoint selection that support good recognition of polygonal models. We have implemented and compared all these measures in a common framework using a user evaluation database to allow for a fair comparison. This framework has been made publicly available and allows to easily include any new measure for comparison, or to use another database as ground-truth. We have also presented a short list of measures that effectively
represent the viewpoint preferences of the users and together with the application fields that the different measures have been employed in.

This contribution has been submitted to ACM Transactions on Applied Perception titled *A survey of viewpoint selection methods for polygonal models.*

- We have quantified in different ways the information associated to the polygons of a 3D model. This information has been used for visualization, viewpoint selection, and object exploration.

  Defining a visibility channel between the polygons of a 3D model and a set of viewpoints, we obtain several shading approaches of the model using the polygonal information. Two of these shading measures are perceptually related to diffuse shading, where image intensity depends on the degree of self-occlusion, and a third measure represents a novel perceptual approach. Several experiments show that the polygonal information measures improve on perceiving shape on similar ambient occlusion measures and that the obtained viewpoint quality measures show a good performance to capture object complexity. Finally, the polygonal information measures are applied to select $N$ best views and to object exploration.

  This contribution has been published in Signal, Image and Video Processing, vol. 7, no. 3, pages 467–478, May 2013 titled *Information measures for object understanding.* [Bonaventura 2013a]

- We have applied our object understanding framework to terrain visualization.

  We have obtained and visualized the polygonal information of a terrain model getting several shading approaches. We have also combined the polygonal information measures and the original terrain texture in order to enhance the perception of the terrain shape. Finally, we have used the viewpoint quality measures to get a set of representative views of the terrain.

  This contribution has been submitted to Computer & Geosciences titled *Information measures for terrain visualization.*

- We have analyzed the use of viewpoint quality measures to measure the similarity between two 3D models.

  We have presented a framework for 3D shape retrieval based on the information channel between the set of viewpoints around a 3D model and the model polygons. From this channel we have derived different similarity measures based on the decomposition of mutual information. We have studied the performance of the sphere of viewpoints as a shape representation of the object.

  This contribution has been published in Computer Animations and Virtual Worlds, vol. 26, no. 2, pages 147–156, 2015 titled *3D shape retrieval using viewpoint information-theoretic measures* [Bonaventura 2015]. This journal publication is an extension of the paper *Viewpoint information-theoretic measures for 3D shape similarity* published in Proceedings of the 12th ACM SIGGRAPH International
8.2. Future Work

The work done during the accomplishment of this thesis can be extended in different ways:

- The viewpoint quality measures obtained from the decomposition of mutual information will be extended with the use of Tsallis-entropy.

- By comparing the viewpoint selection measures, we have obtained a short list of measures that individually behave best. However, if we want to investigate a combination of measures we believe we have to consider not only the ones in the short list but also some of the ones that performed not so well. This is because some of those last measures can contribute identifying different aspects than the ones in the short list. We will analyze, in particular, the combination of $I_1$ (not shortlisted), $I_2$ (shortlisted), and $I_3$ (not shortlisted).

- For the object understanding, we have used Tsallis entropy to sharp or smooth the shading obtained with Shannon entropy. Another generalization of Shannon entropy, called Rényi entropy, will also be used to this purpose.

- We will use the polygonal information to navigate above a 3D terrain model similarly to the exploratory tour presented in Chapter 5.

- For object and terrain understanding we will investigate how the shading of the models using polygonal information can help a human observer to better understand the models.

- In the context of shape retrieval, we have seen that our measures perform far from the state of the art. The results could be improved by optimizing the registration process, using other registration measures than $L_2$ distance, investigating other viewpoint measures and looking for an optimal combination of them.

- To compute the viewpoint quality measures and the polygonal information associated with a 3D model, we have used a sphere of viewpoints around the model. When the models have a clear elongation axis such as a pencil or a terrain model, a different distribution of viewpoints could be more convenient. Thus, we will investigate other viewpoint distributions (e.g., ellipsoid, convex hull, hemisphere) to improve the quality of the captured information.

- Malformed models (incorrect normals or triangles seen from both sides) have an impact on viewpoint selection measures and on shape retrieval. We will study its influence on the different defined measures and we will propose a preprocessing step to alleviate this impact.
• The computation time of the different measures and processes presented in this thesis can be accelerated by using GPU techniques.
Bibliography


VRCAI '13, pages 183–190, New York, NY, USA, 2013. ACM. (Cited on pages 4 and 91.)


Bibliography


