Annotating Images through Adaptation: An Integrated Text Authoring and Illustration Framework

Timo Götzelmann, Marcel Götze, Kamran Ali, Knut Hartmann, Thomas Strothotte
Department of Simulation and Graphics
Otto-von-Guericke University of Magdeburg
Universitätsplatz 2, D-39106 Magdeburg / Germany
{timo, marcel, kamran, knut, tstr}@isg.cs.uni-magdeburg.de

ABSTRACT

This paper presents concepts to support authors illustrating their texts. Our approach incorporates content- and feature-based retrieval techniques in multimedia databases containing 2D images and 3D models. Moreover, we provide tools (i) to adapt the retrieval results to contextual requirements and (ii) to ease their integration into target documents. For 3D models the adaptation comprises aspects of the image composition (i. e., the selection of an appropriate view and the spatial arrangement of visual elements) and the selection of appropriate parameters for the rendering process. In addition, secondary elements (e. g., textual annotations or associated visualizations) are smoothly integrated into adapted 2D or 3D illustrations. These secondary elements reveal details about the semantic content of illustrations and author's communicative intentions. They can ease the retrieval, reuse, and adaptation of illustrations in multimedia databases and are explicitly stored in conjunction with the adapted illustrations.

Moreover, we developed a novel technique to support the mental reconstruction of complex spatial configurations by shape icons. With this illustration technique, shape properties of salient objects can be conveyed using abstract-shaped models. We present retrieval techniques to determine appropriate 3D models to be displayed for shape icons. These shape icons along with the other secondary elements are smoothly integrated into the illustration that can be interactively explored by the user.

Keywords

Text-Authoring, Annotation, 3D Graphics, Interaction

1 Introduction

Authors are often confronted with the challenging task to find appropriate images to illustrate their texts. Even if multimedia databases contain ready-made illustrations, the (i) retrieval and (ii) adaptation of illustrations to contextual requirements is expensive and time consuming. Our approach integrates multimedia retrieval techniques within text authoring tools. By selecting text segments, authors can directly define queries for information retrieval systems. Subsequently, the original documents are enhanced with user-selected illustrations as well as with automatically generated figure captions and references to these figures.

The Journal of WSCG, Vol. 15, ISSN 1213-6964
WSCG'2007, January 29-February 1, 2007
Plzen, Czech Republic.

Human illustrators carefully adjust illustrations to the *communicative function* of the embedding document. This comprises (i) an appropriate *image composition* and (ii) an appropriate *layout of secondary elements* [28] or *annotation layout*. A well-balanced image composition guarantees that graphical objects that are referenced within the text are also visible in the illustration and that their spatial configuration is visualized effectively. The layout of secondary elements includes the selection of graphical objects to be annotated, their content, and their spatial arrangement.

However, illustrations created with a particular goal in mind can rarely be used in contexts different from the original. Therefore, it is often essential to adapt the given illustrations according to new communicative functions and to other external requirements imposed by the layout of parent document. A successful adaptation comprises the selection of appropriate *views* which present characteristic visual features of salient visual objects (image composition) and an appropriate layout of secondary elements.

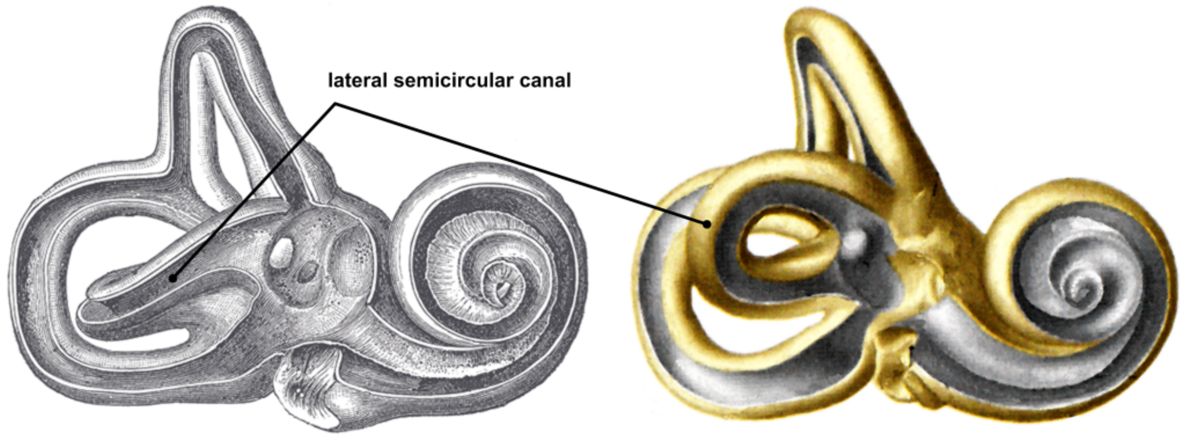

Figure 1: Left: Illustration of the bony labyrinth, an organ of the human ear [11]. **Right:** An altered view to clarify the shape the shape of the annotated object [26].

Layout of Secondary Elements. Textual annotations can establish co-referential relations between textual and visual elements. Therefore, the layout of annotations provides semantic information — detailed descriptions of the content of illustrations. Moreover, it also reflects pragmatic aspects and indication for their communicative function. An explicit representation of annotations both facilitates content-based retrieval techniques and allows an adaptation to different contextual requirements. In this work, we propose to bridge the *semantic gap* of current multimedia retrieval systems by enhancing computer generated images with a formal specification of the layout of annotations.

Image Composition. For some application domains, multimedia databases may also contain computer generated images (e. g., charts, flow diagrams, renditions of surface or volumetric 3D models). Beside standard image processing techniques for 2D illustrations with a restricted potential to adapt the image composition, our visualization component is enhanced to support the adaptation of 3D model renditions to contextual requirements. Human illustrators can interactively select appropriate views and specify textual annotations for visual objects while adaptable real-time algorithms determine annotation layout automatically.

Shape Icons. Regarding the image composition aspect in illustration systems, human illustrators can select only a single point of view to visualize graphical models. But depictions from a single view point neither support the learners to reconstruct mentally the spatial configuration nor they convey characteristic features of all relevant visual objects. In medical education, for example, students have to understand the correct form of objects and have to learn the spatial configuration of their characteristic features. Therefore, anatomic textbooks often contain illustrations of a single object from several viewpoints or illustrators manipulate the spatial configuration in order to present characteristic visual features of the most relevant objects.

We analyzed document variants and found several examples where human illustrators integrated multiple perspectives into a single depiction. Figure 1 presents a correct and a manipulated perspective: The left illustration presents the anatomically correct spatial shape of the *lateral semicircular canal*. Here it is rather difficult to recognize that the canal is shaped like a hollow ring. By rotating the axis of canal in the right illustration, shape recognition becomes easier. However, both illustrations should be presented together in order to convey the correct meaning of the subject.

In order to overcome the limitations of a single visual presentation we developed the concept of *shape icons*. Our idea was inspired by the observation, that even textual description can convey information about the shape or the form of visual objects. In some cases, the object's name itself refers to visual properties or compares the object's shape with well known reference objects (hippocampus–seahorse, cochlea–snail, etc.). Therefore, we implemented a novel tool which suggests shape icons for the most relevant objects in the current interaction context. In order to clarify the three-dimensional form without altering the real spatial configurations, illustrators, instructors, or learners can select an appropriate icon which is then displayed in a textual annotation. We employ shape similarity measures to determine the most relevant one within a small set of visual reference objects.

This paper is organized as follows: Sec. 2 reviews the related work. The architecture of our experimental application is presented in Sec. 3. Sec. 4 describes several application scenarios of our framework. Then the layout of the textual and visual annotations (Sec. 5) and the determination of shape icons (Sec. 6) are explained. Finally, Sec. 7 summarizes our contributions and Sec. 8 discusses some directions of future work.

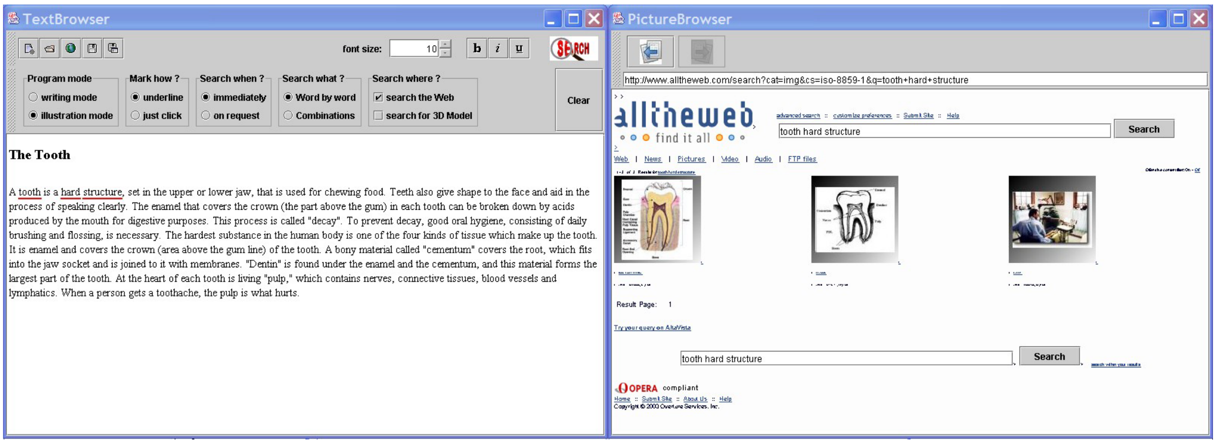

Figure 2: The SearchIllustrator [12].

2 Related Work

This section aims at giving a short overview of techniques to retrieve 2D illustrations or 3D models from multimedia databases. Additionally, a brief survey of methods to adapt the rendition of 3D models according to communicative goals, and layout scheme of secondary elements are presented.

Image Retrieval Techniques. Due to the availability of comprehensive multimedia databases and content-based retrieval techniques, the text illustration process has shifted from content creation to search with respect to communicative goals. The strategies of experienced practitioners as well as their advantages and disadvantages have been described by the journalists Markkula and Sormunen [21]. They report that journalists employ a keyword-based search in huge image databases. Since these keywords might not match exactly the manually created image descriptions, they are very often afraid of missing an appropriate image and therefore tend to create queries that produce many results. The process of browsing through the results, the manual insertion of the search results into the text may become very time consuming and takes a big part of the effort required for creating an illustration. To ease this burden, our system automatically inserts illustrations into the text and generates initial figure captions on the basis of the query (see Fig. 2).

The automatic retrieval of multimedia content (such as images or 3D models) is a relatively young and highly competitive research area, where descriptions of the image's content are either extracted from the data itself (e. g., color histograms or distributions in 2D, shape characteristics in 2D and 3D) or from contextual information and manual annotations (*metadata*). Retrieval techniques which extract features from the data do not require manual annotations; however, they do not support content-based queries (the so-called *semantic gap*).

Manually created descriptions of the image's content are often incomplete, inconsistent, and language dependent. Moreover, the pure amount of images to be annotated raises severe problems for image retrieval systems. Therefor good search engines incorporate *collaborative* or *social tagging* approaches [29, 30] to consistently annotate the semantic content of a huge amount images which can be found on the WWW.

Liebermann and Liu present an interesting approach that shows how image retrieval can benefit from the analysis of semantical relations between concepts. The authors present a system that analyzes annotations in images and uses world semantics to make image retrieval more robust [20]. Another approach relies on statistical analysis of relevance feedback [19]. It uses a post processing step to improve the retrieval performance in a way that more semantically-related images are returned.

3D Retrieval Techniques. The search for 3D models is based on different similarity measures. It employs spatial (shape) distributions of vertices in 3D models [22], symmetry axis [17] or skeleton graphs [5], approximations of complex shapes with sets of simple geometric objects [27], or transformations of 3D models into frequency representations [18]. These approaches can be refined by iterative user feedback mechanisms [9]. Some engines offer a web interface and present their results in a browser window. Moreover, the Princeton 3D model search engine also allows users to sketch the shape of the desired objects or to search for text linked with 3D models [10].

Interactive Illustration Techniques. Computer generated renditions of 3D models can automatically be adapted to emphasize the most salient objects in a document to be illustrated [15]. Due to (partial) occlusions, a single illustration often does not suffice to depict all salient visual objects. The illustrative browser [25] restricts the number of salient objects to those contained in the current displayed text segment

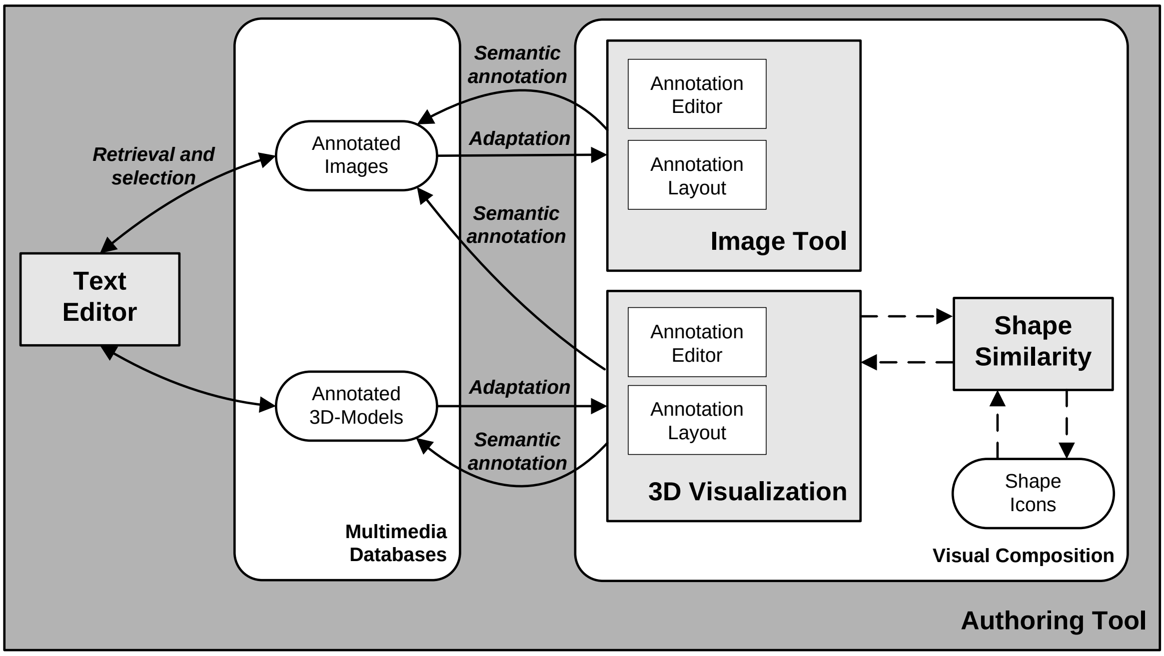

Figure 3: System overview.

while non-photorealistic rendering techniques can also present occluded objects. Moreover, learners can interact with the visualization and with the textual part of the document. Users can change the view of graphical objects in the illustration and the textual part is scrolled accordingly to show the matching explanations and vice versa. However, this is only possible in an interactive environment.

Layout Algorithms. The research on the layout of annotations was pioneered by the cartographic community [16]. There exist a wide variety of research prototypes to integrate annotations into interactive information systems such as dynamic maps [23] and medical and technical illustrations [1, 13, 6]. Recently, the term view management was introduced in Augmented and Virtual Reality for a more general, but related problem: the smooth integration of additional 2D information (images, texts, annotations) into the view plane [3, 2].

3 Architecture

Our approach extends the SearchIllustrator concept [12] that employs information retrieval techniques on multimedia databases or web search engines to interactively illustrate texts (cf. Fig. 2). The search can be performed in two ways. First, the user can interactively select keywords that control a search engine for static images and 3D models. Second, the system analyzes the text and performs a background search during the writing process. After the creation of the text is finished the system presents a collection of possible images or 3D models for illustration. User selected images are not adjusted to contextual requirements, whereas the parameters for viewing direction and the rendering style are adjustable. The 3D model is then used to create different photorealistic, non-photorealistic, and hybrid renditions, depending on the users needs and the communicative goal.

Within the *authoring tool* (see Fig. 3), illustrators can directly access the results of a multimedia retrieval system and select appropriate images or 3D models. Subsequently, an interactive *3D visualization* system allows to adapt the viewing direction and the rendering style to contextual requirements. Our implementation extends Götze's [12] original framework with a flexible real-time annotation system. In an *annotation editor*, illustrators can specify textual annotations for visual objects and adjust their placement. An automated *annotation layout* system determines a frame-coherent layout during user interactions which considers and retains manual specifications of annotations.

The *shape similarity* module assists users to determine shape icons for relevant 3D objects. The system suggests a ranked list of candidate 3D objects, while the user selects an appropriate model and adjusts the viewing direction. Subsequently, small projections of these 3D models (*shape icons*) are integrated into the layout.

The link between the 2D and 3D visualization system highlights the fact that our approach extends computer generated renditions with explicit specifications of the rendering parameters and the annotation layout, so that illustrators or readers may access the underlying 3D visualization through computer generated images contained in interactive documents.

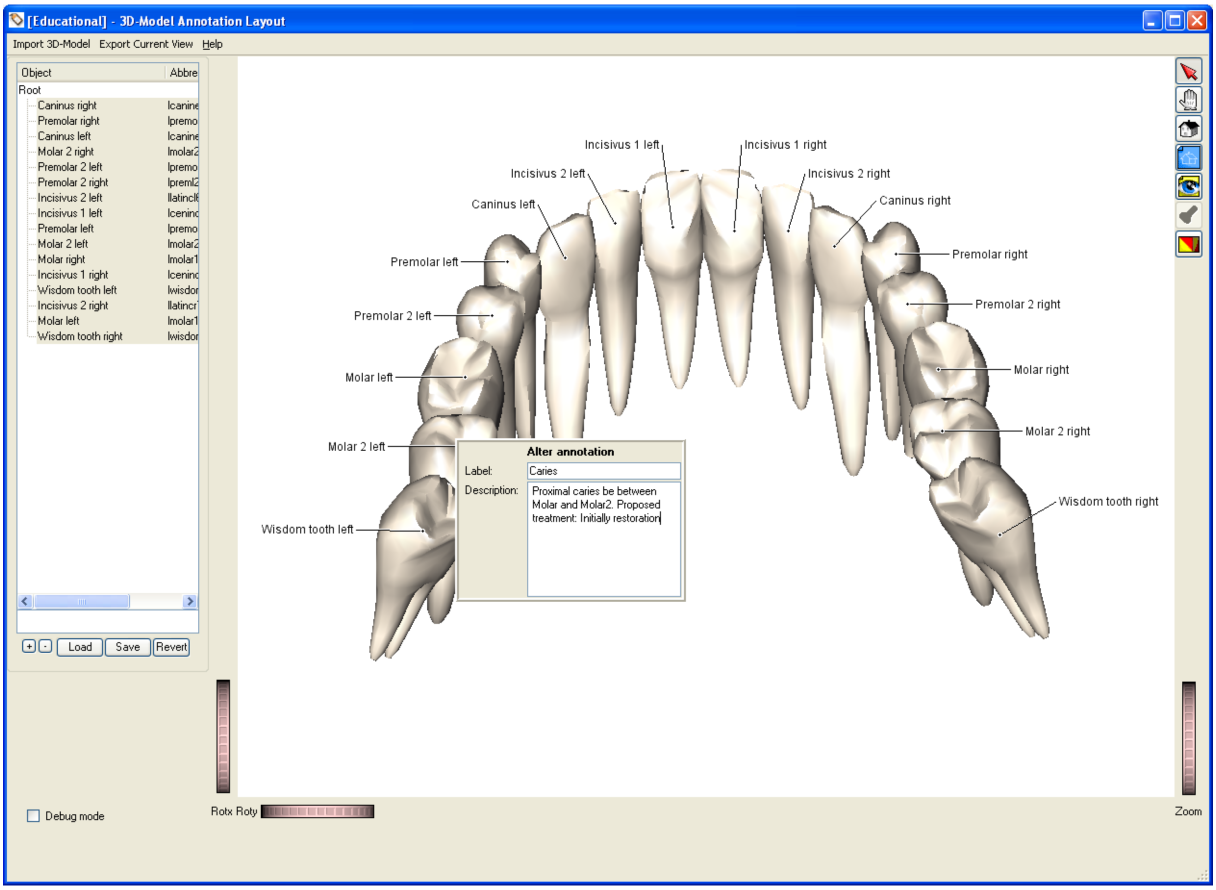

Figure 4: Interactive annotation of 3D models.

4 Scenarios

This section describes 3 different application scenarios of our approach by referring to the several modules of the architecture. In each of the scenarios the starting point is a text to be illustrated in the *text editor*. In first scenario (cf. Sec. 4.1), system looks for appropriate 2D images in *multimedia database* which are then integrated into the visualization. In the second scenario (cf. Sec .4.2), the system adapts the relevant 3D model to contextual requirements of the text and stores the *semantical annotations* explicitly as text in the resulting illustration. The last scenario (cf. Sec. 4.3) inserts *shape icons* in the visualization in order to avoid ambiguity in spatial details.

4.1 Adequate Illustration

Let's suppose, a user wants to illustrate a text of a document with an adequate image. He/she marks some of the terms used in the text and submits the query to the retrieval module. The module displays some searched illustrations, one of which can be chosen by the user and integrated into the text via a mouse click. Here, no adaptation of the illustration is performed by the system.

4.2 Adapted Illustration

The second scenario assumes that there is a user who wants to illustrate a text in a specific context. To search for an appropriate illustration, he/she marks the relevant terms in the text editor. The retrieval module offers several different search results. The user selects one of the 3D models that can be used to roughly illustrate the text. When the view and the annotations of the visualization do not optimally correspond with the context described in the text, the user utilizes the 3D visualization module to interactively choose an appropriate view of the 3D model and adapts the annotations to the text's contextual requirements. Finally, the adapted 3D model is integrated into the text editor and saved for later use. Since semantical annotations describe the content of illustrations, our approach stores annotations explicitly in a textual fashion, hence, the retrieval system can use them for future searches to regain and re-use them.

4.3 Illustration with Shape Icons

In this scenario, a user needs a special illustration related to a very specific context. Therefore, it is required to adapt the illustration according to the corresponding text into the text browser. Within our system,

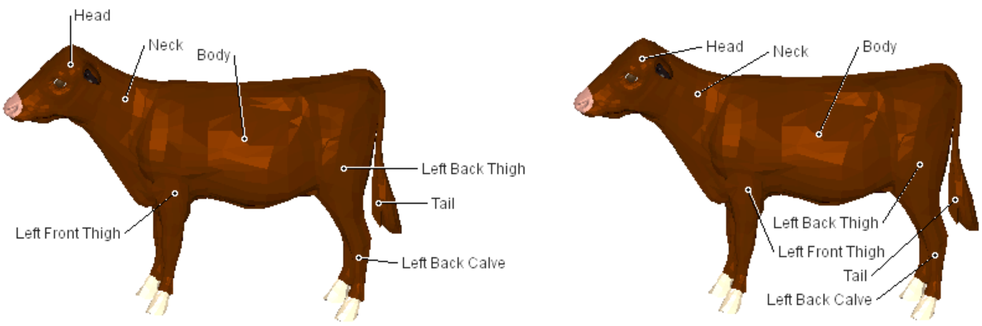

Figure 5: Left: A suggested annotation layout. **Right:** Integration of manual layout constraints to meet external layout restrictions.

the user can change the annotations and the view of a 3D model retrieved by the search module. However, it might not always be possible to find a view which shows the spatial extents of all important 3D components in an unambiguous way. To solve that problem the user involves the shape similarity module to retrieve shape icons which help to disambiguate the spatial shape of the objects.

5 Annotation Layout

The adaptation of retrieved visual material to new contextual requirements comprises their (re)composition and the enhancement with additional information. The determination of an appropriate viewing direction for a 3D model or the selection of a display window for a 2D illustration involves semantic, pragmatic, and aesthetic considerations which should be done by a human expert (see Blanz's [4] psychological experiments to determine canonical views and Polonsky's [24] review of algorithms to determine "good" views of three-dimensional models.). In contrast, there are good heuristics for a functional and aesthetic layout of annotation [16, 8, 14]. Therefore, we developed tools which support authors to adjust the visual composition for 2D illustrations as well as for computer generated renditions of 3D models and to add additional information to visual elements or alter their content (see Fig. 4). An automated layout system determines the placement of all annotations and considers constraints posed by the illustrator.

The automatic layout of annotations considers the spatial configuration on projection in real-time. We incorporate a potential field approach on color-coded projections [13]. The novel contributions of this paper are an *annotation editor* and the integration of manual constraints into the automatic *annotation layout* (see Fig. 4 and 5). Illustrators can define annotations by selecting arbitrary positions on the image (2D) or on the surface of the 3D models. Moreover, their content can be altered by selecting the desired visual object or its annotation. Finally, the layout of the target document often imposes restrictions on the maximal size of an embedded illustration, which heavily influences the layout of annotations. In order to allow the user to correct unaesthetic placements, manual layout specifications are considered in the layout algorithms (see Fig. 5).

6 Determination of Shape Icons

In order to support the mental reconstruction of complex spatial configurations, instructors and learners can add images of similar 3D objects. Our system employs shape similarities to suggest an appropriate object as a *shape icon* from a predefined set of 3D reference objects. Moreover, the object itself is included in this list, as it might be partially occluded or depicted from a non-canonical view. Finally, the texts presented in annotation can be used to for queries for image or 3D model search engines. The three most similar reference objects and the object itself are displayed. After choosing a shape icon, it is accordingly displayed next to the textual annotation (see Fig. 6).

We integrated Chen's 3D retrieval engine [7] because one could specify a corpus of 3D reference objects. Of course, it is possible to use other search engines instead.

The sequence of determining spatially similar objects is as follows: In a pre-computation step, shape de-

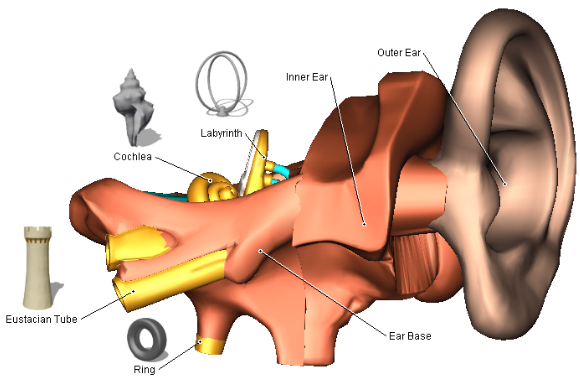

Figure 6: An annotated ear with several shape icons to disambiguate the spatial shape of specific objects.

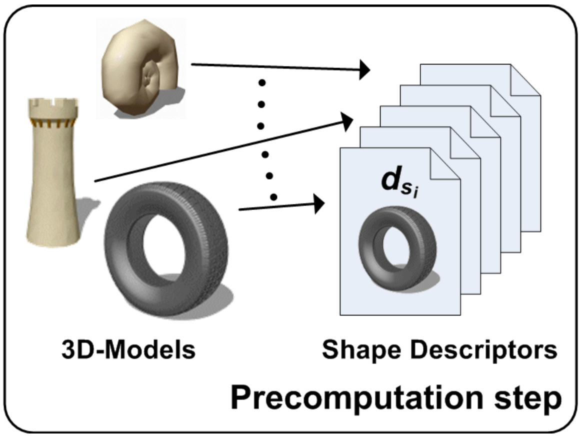

Figure 7: Preprocessing step.

scriptors for each object in the database of reference objects are determined (see Fig. 7). If the author is selecting a specific object of the 3D model and requests a shape icon, accordingly a new shape descriptor for this object is determined, too. Subsequently, the system computes the similarity of the selected objects shape descriptor with each of the pre-computed shape descriptors of the reference objects in the database (see Fig. 8).

Additionally, the textual annotations associated with objects are used for keyword-searching. Next, the search results are ranked and the candidates are presented according to their score.

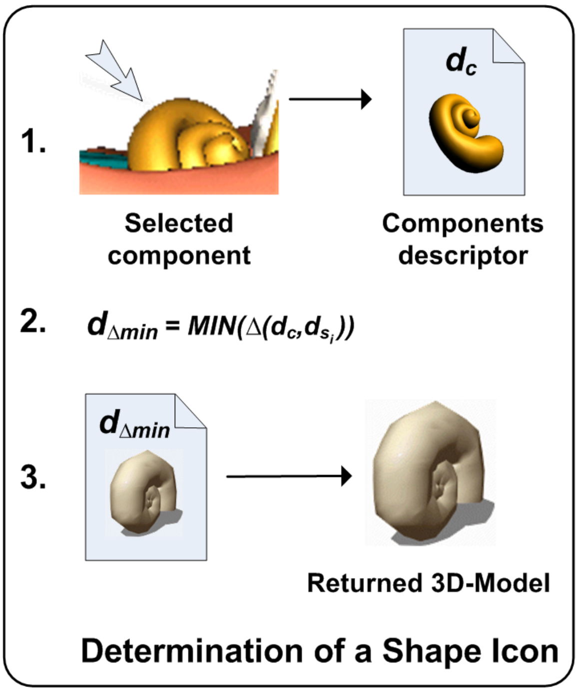

Figure 8: Shape matching step.

By selecting one of the candidates, the author can adopt it as a shape icon. Finally, the annotation layout is recomputed.

To render a shape icon from 3D objects, the set of reference 3D objects also contains specifications of canonical views. Another approach is to align two canonical directions (front and top) between the selected and the reference 3D object, and to adjust the view of the shape icon to the current viewing direction. To determine the appropriate strategies, however, user tests are required.

7 Conclusion

In this paper, we developed a novel concept to support the interactive illustration of texts with content-based search strategies in multimedia databases. The main contributions are: (i) We proposed a new kind of interactive documents by retaining the rendering parameters for computer-generated projections so that readers can directly access 3D visualizations of complex spatial configurations. (ii) The definition of textual annotations for visual objects and their appealing and frame-coherent presentation in interactive 3D visualizations and 2D illustrations is a central element of the adaptation of predefined visual materials to contextual requirements. Our approach considers the annotation layout as an inherent description of the semantic and pragmatic content of illustrations. Hence, their explicit representation eases content-based retrieval techniques and the reuse and adaptation of images. (iii) We introduced the concept of shape icons to clarify renditions of complex spatial shapes. Appropriate geometric reference objects are determined by a combination of shape and keyword-based 3D retrieval techniques and are interactively selected by instructors or learners in order to ease their mental reconstruction. (iv) We implemented an experimental application which offers all basic functionalities.

8 Future Work

Since this framework is designed in a modular fashion, it is possible to integrate additional modules to it which aid the illustrator to emphasize several parts of the illustrations. To ensure the visibility of all important parts of an 3D object, human illustrators often use visual techniques like transparency (ghosting) and cutaways. Thus, we are currently investigating a set of those techniques.

Though, the discussions with anatomists revealed that shape icons could improve medical training, they have to be evaluated. Thus, we plan a user study to evaluate our system. Some tests could compare the effectivity of unchanged illustrations found in the WWW with those which were adapted via our system. Another test could reveal the time efficiency of our integrated approach compared with a manual search and adaptation of appropriate illustrations.

References

[1] K. Ali, K. Hartmann, and T. Strothotte. Label Layout for Interactive 3D Illustrations. *Journal of the WSCG*, 13:1–8, 2005.

[2] R. Azuma and C. Furmanski. Evaluating Label Placement for Augmented Reality View Management. In *IEEE and ACM Int. Symp. on Mixed and Augmented Reality*, pages 66–75, 2003.

[3] B. Bell, S. Feiner, and T. Höllerer. View Management for Virtual and Augmented Reality. In *Symp. on User Interface Software and Technology*, pages 101–110, 2001.

[4] V. Blanz, M. J. Tarr, and H. H. Bülthoff. What Object Attributes Determine Canonical Views? *Perception*, 28:575–599, 1999.

[5] A. Brennecke and T. Isenberg. 3D Shape Matching Using Skeleton Graphs. In *Simulation und Visualisierung*, pages 299–310, 2004.

[6] S. Bruckner and E. Gröller. VolumeShop: An Interactive System for Direct Volume Illustrations. In *IEEE Visualization*, pages 671–678, 2005.

[7] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung. On Visual Similarity Based 3D Model Retrieval. *Computer Graphics Forum*, 22(3):223–232, 2003.

[8] S. Edmondson, J. Christensen, J. Marks, and S. Shieber. A General Cartographic Labeling Algorithm. *Cartographica*, 33(4):13–23, 1997.

[9] M. Elad, A. Tal, and S. Ar. Content based Retrieval of VRML Objects — An Iterative and Interactive Approach. In *EG Workshop in Multimedia*, pages 107–118, 2001.

[10] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, and D. Jacobs. A Search Engine for 3D Models. *ACM Transactions on Graphics*, 22(1):83–105, 2003.

[11] H. Gray. *Anatomy of the Human Body*. Lea & Febiger, Philadelphia, 20th edition, 1918.

[12] M. Götze, P. Neumann, and T. Isenberg. User-Supported Interactive Illustration of Text. In *Simulation und Visualisierung*, pages 195–206, 2005.

[13] T. Götzelmann, K. Hartmann, and T. Strothotte. Agents-Based Annotation of Interactive 3D Visualizations. In *6th Int. Symp. on Smart Graphics*, pages 24–35, 2006.

[14] K. Hartmann, T. Götzelmann, K. Ali, and T. Strothotte. Metrics for Functional and Aesthetic Label Layouts. In *5th Int. Symp. on Smart Graphics*, pages 115–126, 2005.

[15] K. Hartmann and T. Strothotte. A Spreading Activation Approach to Text Illustration. In *2nd Int. Symp. on Smart Graphics*, pages 39–46, 2002.

[16] E. Imhof. Positioning Names on Maps. *The American Cartographer*, 2(2):128–144, 1975.

[17] M. Kazhdan, B. Chazelle, D. Dobkin, T. Funkhouser, and S. Rusinkiewicz. A Reflective Symmetry Descriptor for 3D Models. *Algorithmica*, 38(2):201–225, 2003.

[18] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. In *Symposium on Geometry Processing*, 2003.

[19] M. Li, Z. Chen, and H. Zhang. Statistical Correlation Analysis in Image Retrieval. *Pattern Recognition*, pages 2687–2693, 2002.

[20] H. Lieberman and H. Liu. Adaptive Linking Between Text and Photos using Common Sense Reasoning. In *2nd Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 2–11, 2002.

[21] M. Markkula and E. Sormunen. Searching for Photos — Journalists' Practices in Pictorial IR. In *The Challenge of Image Retrieval, A Workshop and Symposium on Image Retrieval*, 1998.

[22] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape Distributions. *ACM Transactions on Graphics*, 21(4):807–832, 2002.

[23] I. Petzold, G. Gröger, and L. Plümer. Fast Screen Map Labeling — Data Structures and Algorithms. In *21st Int. Cartographic Conf.*, 2003.

[24] O. Polonsky, G. Patané, S. Biasotti, C. Gotsman, and M. Spagnuolo. What's in an Image? *The Visual Computer*, 21(8–10):840–847, 2005.

[25] S. Schlechtweg and T. Strothotte. Illustrative Browsing: A New Method of Browsing in Long On-line Texts. In *Int. Conf. on Computer Human Interaction (INTERACT-99)*, pages 466–473, 1999.

[26] J. Sobotta, R. Putz, and R. Pabst, editors. *Sobotta: Atlas of Human Anatomy*. Lippincott Williams & Wilkins, Baltimure, 13. edition, 2001.

[27] M. Suzuki. A Dynamic Programming Approach to Search Similar Portions of 3D Models. *The World Scientific Engineering Academy and Society Transaction on Systems*, 3(1):125–132, 2004.

[28] E. R. Tufte. *Visual Explanations: Images and Quantitatives, Evidence and Narrative*. Graphics Press, Cheshire, Connecticut, 1997.

[29] L. von Ahn and L. Dabbish. Labeling Images with a Computer Game. In *SIGCHI Conf. on Human Factors in Computing Systems*, pages 319–326, 2004.

[30] L. von Ahn, R. Liu, and M. Blum. Peekaboom: A Game for Locating Objects in Images. In *SIGCHI Conf. on Human Factors in Computing Systems*, pages 55–64, 2006.