1 Introduction

Despite great advances of computer graphics capabilities, rendered images are rarely used for educational purposes. It is still common practice to use hand-drawn illustrations in textbooks (see [34] for a detailed discussion). A variety of illustration techniques are employed: objects are not always drawn entirely to scale, the level of detail varies to a great extent over the image, objects are often drawn in unnatural colors. Therefore, we conclude that abstraction techniques are essential to enhance computer generated illustrations with respect to educational purposes. Illustrations in textbooks are accompanied by figure captions which describe illustration techniques.

The orientation provided by such captions is well-recognized in psychology. GOMBRICH, for instance, argues that "No picture tells its own story" [10]. WEIDENMANN considers this statement to hold also for educational situations [37].

Advances in computer graphics make it possible to influence graphics generation on different levels and thus to adapt visualizations to the information extraction task of the user. For example, within the ZOOM-ILLUSTRATOR [23] a number of abstraction techniques for an interactive exploration of anatomic models are incorporated (see discussion of abstractions in the ZOOM-ILLUSTRATOR in Section 3.1). As a result of the application of abstraction techniques, the illustration may not correspond to a physically correct image of the depicted model, hence these illustrations may be misinterpreted. Based on this observation, we furnished the interactive illustrations of the ZOOM-ILLUSTRATOR with figure captions.

In this framework, figure captions serve at least two different purposes: First and foremost, figure captions describe what is depicted. Moreover, the effects of abstraction techniques are described (what has been removed or scaled up or down). Therefore, figure captions may guide the interpretation of such images. In this paper some aspects of the automated generation of figure captions are described.

The paper is organized as follows: First, some basic definitions and classifications are provided in Section 2. First, the term abstraction is defined. Based on this definition, typical abstraction techniques are briefly surveyed. Second, we classify figure captions according to their function.

In Section 3 we describe the user interactions and abstractions in the ZOOM-ILLUSTRATOR system and incorporation of figure captions in this anatomic tutoring system. The incorporation of figure captions in interactive systems raises the issue of describing visualizations exposed to changes. Different strategies for the update of figure captions due to changes of the visualization are described. The content of these figure captions can be adapted to user preferences. An entirely new concept is to use figure captions as input for the graphics generation process.

In Section 4, we propose a framework in which abstraction is automatically or interactively applied to generate visualizations of information spaces together with figure captions describing this visualization. Moreover, a prototypical implementation using template-based generation is presented. We conclude with the discussion of related work (Section 5), of open questions of the current approach (Section 6) and a summary (Section 7).

2 Abstraction and Their Description in Figure Captions

In this section basic terms such as abstraction are defined. Furthermore, abstraction techniques are classified (Section 2.1). Finally, a classification of figure captions according to their content is provided (Section 2.2).

2.1 Abstraction in Illustrations

Illustrations are produced to enable a viewer to extract information. For this purpose, they contain not merely straightforward renditions of the data—the presentation is dedicated to a *thematic focus*. Thus, portions of the data may be presented in more detail or more comprehensively, others may be simplified, shrunken or even left out.

We refer to the data from which an image is generated as an *information space*. We use the term *abstraction* to denote "the process by which an extract of a [...] information space is refined so as to reflect the importance of the features of the underlying model for the dialog context and the visualization goal at hand" [35, p. 14].

Abstraction introduces a *distortion* to the visualization with respect to the underlying model. The *fidelity* of a visualization depends on the kind and degree of abstraction applied to single objects or classes of objects.

For computer generated graphics, data on the fidelity can be obtained as a direct by-product of the visualization process. Thus, the abstraction can be evaluated with respect to the inaccuracy and with respect to how the abstraction violates the expectations of the intended user. The assessment of the user "expectations" is a difficult task and, of course, requires an intensive study of the different information exploration tasks in a given application domain, including heuristic or empirical evaluations. In anatomy, for example, it is important for a student to recognize distortions of topological relations and relative sizes.

Abstraction techniques are the means by which the effect of the abstraction process is achieved. Since there are usually several abstraction techniques which produce a similar effect, the designer of an illustration has to select one or a combination of several abstraction techniques. To provide visual access to an object, for example, the objects occluding it may be removed or relocated, a cut-away view can be used or a rotation of the model can be employed. This choice is constrained by parameters of the output medium chosen and of human recognition.

To meet the restrictions of human recognition it is essential to establish an equilibrium between the level of detail of the objects of interest and the objects depicting their context so that the user can understand the image. On the one hand, it is crucial to reduce the cognitive load for the interpretation of an illustration. On the other hand, enough contextual information must be provided to understand an image.

In anatomy, for example, bones are important landmarks for the localization of muscles and ligaments. Even in an image focusing on ligaments, parts of the skeleton should be depicted to enable the viewer to recognize the spatial location of the ligaments. Human designers frequently resolve this conflict by drawing objects within the thematic focus and the context in different styles (they often reduce the attraction of unfocused objects by illustrating them grey, with reduced level of detail, or only as a silhouette).

Inspired by observations from hand-drawn illustrations, similar techniques for the generation of rendered images have been applied to graphics generation systems, like [8], [32] and [31]. These techniques work on different levels: **high-level abstraction techniques** are concerned with *what* should be visible and recognizable. Above all, the viewing specification determines the visible portion. Furthermore, the most salient objects should be centered. A combination of several images could be used if the objects of the thematic focus cannot be depicted from one view point. Abstraction techniques such as cut-aways, exploded views, simplification, zooming (for a discussion of fish-eye zoom techniques see Section 3.1) ensure the visibility and discriminating power of the objects within the thematic focus. Lighting specification influences the composition and thus the content selection for the graphics generation.

Low-level abstraction techniques, on the other hand, deal with *how* objects should be presented. Colors, textures, brightness values as well as line styles and cross-hatching techniques are frequently adapted to support the visualization goal. We refer to these as *presentation variables* (see also [20]). However, as low-level abstraction techniques also determine the contrast of adjacent objects and the contrast of objects to the background, they influence the composition of the illustration, too.

Since visualization is a sophisticated process, manipulations that restrict image fidelity ought to be described in the caption. This covers the mention of single objects whose visibility or position was adapted during the visualization process. If users shall be made aware of these modifications, the objects need to be described in a figure caption.

2.2 Classification of Figure Captions

BERNARD [2] introduces the terms descriptive and instructive figure captions. According to BERNARD, **descriptive figure captions** provide a verbal description of the content of an image. Our analysis of figure captions in anatomy suggests using this definition in a broader sense. We use the term *descriptive figure captions* to refer to a description of a view on a model or a section of the real world. This new definition also comprises phrases describing not only what is visible in the figure but also what is hidden, has been removed, or what important objects are close to the depicted portion. Consequently, in our terms descriptive figure captions describe an image and its (spatial) context.

Descriptive figure captions serve various functions: They summarize the content of pictures and ease the orientation by providing the name of the depicted model and the viewing direction. Furthermore, descriptive figure captions often inform about the thematic focus and describe abstraction techniques applied to an image. Since these modifications often remain unnoticed to the untrained eye, their existence needs to be described. This is probably the most interesting aspect of the figure captions' functionality, as it reveals what has happened in the abstraction process. These parts reflect the abstraction process, i.e., the operations that have been performed on the data to obtain the visualization, and their impact on the fidelity of the visualization.

Instructive figure captions explain how an action could be performed. These actions are often denoted by meta-objects, i.e., graphical objects like arrows which "do not directly correspond to physical objects in the world being illustrated" [32, p. 127] in the accompanying illustration. These captions are often employed to describe several stages within complex actions. Despite of being inspired by the term itself, our definition clearly differs from BERNARD's usage, where instructive figure captions are intended to focus the attention of a viewer on important parts of the illustration.

If an illustration depicts certain stages within an action or process, it represents an abstraction over time. The meta-objects guide the mental reconstruction of the process which leads to the depicted stage or which starts from the depicted stage. Instructive figure captions can provide information over and above the depicted content, e.g., causal relationship of steps within a complex action, pre- and postconditions of actions, and the instruments needed to perform this action. Furthermore, instructive figure captions can mention that the depicted stage presents an unwanted situation which arises from a typical complication. Instructive figure captions can be found, for example, in technical documentation and textbooks on surgery. However, we will not refer to such instructive figure captions in more detail.

This classification covers a large portion of captions in reference materials, textbooks and repair manuals. Other figure captions, e.g., citations of people shown on a photograph in a journal, are beyond the scope of this classification.

3 Figure Captions for Interactive Anatomical Illustrations

This section presents which kind of graphical manipulations (abstractions) can be performed by the ZOOM-ILLUSTRATOR system (Section 3.1). Figure captions in interactive systems are exposed to various changes. To be consistent with the image they describe, they have to reflect these changes. However, if every interaction with an image leads to a complete replacement of the figure captions this is highly irritating. Therefore it is an essential issue when figure captions are updated and who initiates this process. Figure captions may change in content and length. This requires a flexible layout of the overall illustration in which figure captions may "grow". We therefore refer to them as *dynamic figure captions* (Section 3.2). Furthermore, the structure and the content of figure captions should be adjusted in order to meet user's preferences. Those *adaptable figure captions* are described in Section 3.3. Moreover, *interactive figure captions* provide means to adjust the visualization. New challenges and problems which arise with that concept are discussed in Section 3.4.

3.1 ZOOM-ILLUSTRATOR

The ZOOM-ILLUSTRATOR presents one or two views onto an object, together with labels which contain annotations. There may be several annotations of an object in different levels of detail. Whenever the user selects an object or its annotation, more detailed object descriptions are presented within the accompanying label, whereas the amount of text in other labels or the number of labels has to be decreased. So the varying amount of text presented in one annotation enforces a rearrangement of all labels.

This is accomplished by a variant of the *continuously variable zoom* as introduced by DILL et al. [7]. The variable zoom is a fish-eye zoom algorithm which manipulates rectangular areas, called nodes. One node may be enlarged at the expense of others the size of which are reduced accordingly.

Figure 1: *A caption has sensitive parts (with an underlying rectangle) and invalid parts (light grey and italic). Bold parts present fixed phrases which are based on the templates.*

This non-linear scaling is inspired by fish-eye lenses used in photography which result in a large magnification of central regions where as the peripheral regions are reduced in size. While the original zoom algorithm works on rectangular areas in two dimensions Raab and Rüger [28] extended it to the manipulation of arbitrary-shaped objects in 3D. This fish-eye zoom preserves the topological relations between all objects in 2D or in 3D whenever the size of objects is changed. One effect of this algorithm is to preserve contextual information while changing the focus of an illustration.

As a consequence of the request for additional information about an object is the application of several techniques of graphical emphasis. Emphasized objects should be clearly visible and recognizable. This can be achieved by altering presentation variables of the object itself or those of other objects. Frequently, those objects occluding the object to be emphasized are drawn semi-transparent, while the saturation of the object to be emphasized is increased.

Furthermore, geometric manipulations, such as the 3D zoom, may be applied to highlight objects. Finally, the system may determine another viewing position using some heuristics, for instance to increase the visible portion of the object to be highlighted.

There is an [online description][1] of the Zoom-Illustrator system available as well as a [gallery of screen-shots][2],which may be useful to illustrate the ideas presented above. Furthermore, several papers ([25], [24], [23]) discuss various aspects of the Zoom-Illustrator system in more detail.

3.2 Dynamic Figure Captions for Interactive Systems

We refer to those parts of the figure caption that do not reflect the current illustration correctly as *invalid*. In interactive systems like the Zoom-Illustrator, figure captions describe images exposed to changes. In practice, nearly all interaction on images leads to incomplete or even invalid figure captions.

Hence, invalid parts must be detected automatically. Furthermore, these parts must either be updated or their status has to be communicated by means of an appropriate layout (see Figure 1). This raises the question of when the caption is to be updated and who initiates the update process. We use three approaches:

Update on explicit request only. The caption is updated only upon the user's request. The system indicates invalid parts.

Automatic update. Compared to the other variants, automatic updates by the system require fast generation of captions. The high frequency of

[1] Hyperlink: http://www.cevis.uni-bremen.de/ bernhard/eng/zoomIllu.html
[2] Hyperlink: http://www.cevis.uni-bremen.de/ bernhard/ger/gallerie.html

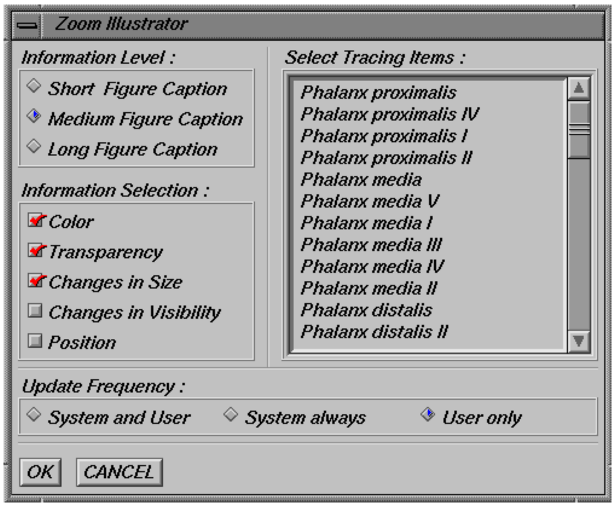

Figure 2: *Adaptation of figure captions. In the left part, the content selection is customized. In the right part, the user selects objects for monitoring.*

changes in the corresponding area of the screen may confuse users. However, the advantage of this approach is that captions remain consistent with the image at all times.

Hybrid variant. The caption is updated automatically only if radical changes in the image occur. Examples are considerable changes in the viewing specification (e.g., via a rotation) leading to a change in the visibility of a large number of objects or the removal of a class of objects due to filtering operations. Another example is the incorporation of an additional view which requires the figure caption to describe both views.

As there is no optimal variant for all users, it seems to preferable to leave the decision on the initiation of an update up to the user (for the update frequency option see the bottom line in Figure 2).

Another aspect of variation concerns the fact that figure captions in interactive systems may change in length over time. The incremental completion of an existing figure caption where constant parts remain at their positions facilitates the identification of changing parts. This strategy may, however, conflict with a well-balanced layout where captions are centered below an image.

3.3 Adaptable Figure Captions

Another motivation for using dynamic figure captions is to tailor the content of these captions to the user's preferences. Such an adaptation is essential because the overall amount of information mentioned in a figure caption can become quite large. Adaptation facilities should enable users to control the content selection. Several parameters are particularly important for adaptation:

Content selection. The caption generation process can be adapted with regard to what kind of information is to be presented and to the level of detail. The information selection option (see the mid-left part of Figure 2)

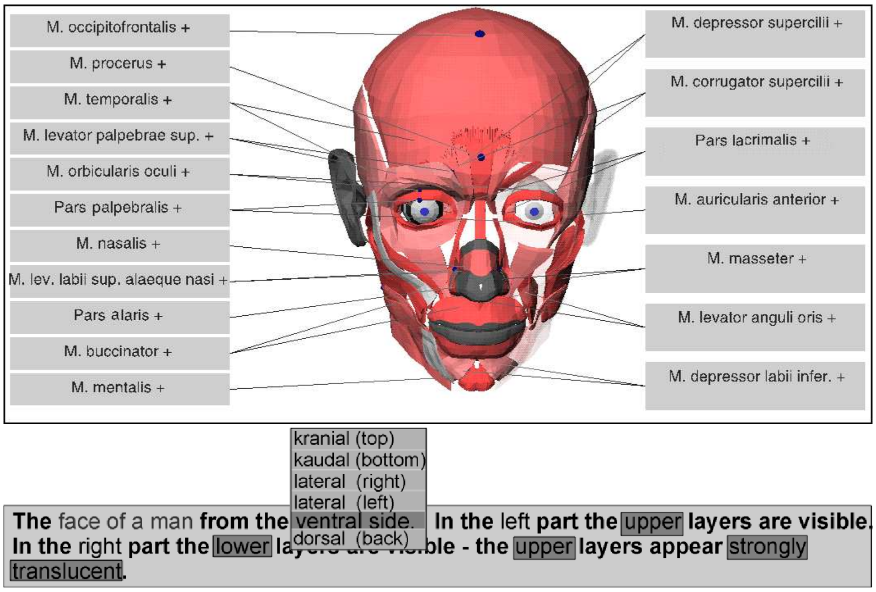

Figure 3: *Interactive figure captions. The selection of "lateral" causes a pop-up menu to appear containing six main viewing directions.*

allows a user to request a notification when a presentation variable has been altered, or when geometrical aspects of the image have been manipulated.

In addition, a user may control the amount of information by specifying the level of detail (see the information level option in the top left part of Figure 2). Selecting short figure captions means that only the most important changes in each selected category are presented, whereas long figure captions inform the user about every modification to the selected property.

Object monitoring. Users should be able to express their interest in specific objects or regions (parts of the underlying model). Based on such a specification, the user will be informed about all changes that affect the objects or regions traced and about their current state (see the *select tracing items* option in the right part of Figure 2). For instance, if an object becomes hidden, the system comments on the visibility of such objects. This is similar to debugging tools which allow the user to monitor variables.

3.4 Interactive Figure Captions

Figure captions may contain sensitive parts which can be manipulated and cause an update of the image. With this approach figure captions serve as input for the graphics generation process. Therefore, we refer to these captions as *interactive* figure captions.

Usually, an image can be manipulated with very different handles. There are dialogues or handles to invoke transformations, material editors to change presentation variables and other dialogues to initiate filtering operations. The interactive manipulation of the corresponding parts of a figure caption, now, unifies these interaction facilities.

Interactive figure captions are also attractive in another respect—a figure caption does not only describe an image but also makes it possible to modify it. The image is no longer superior to the caption in that the caption is dependent on the image, but both stand on the same level and offer interaction facilities.

It is necessary to guide the user through this interaction. An unrestricted editing of the caption may lead to the specification of infeasible or ambiguous commands. Moreover, the user may not be aware of what interaction options are available. Guiding the user also avoids the enormous problems coming along with natural language understanding. Therefore, sensitive regions are provided (see Figure 1) which display simple dialogues with alternative choices for this item when selected. These can be pop-up menus, sliders, text edit fields for the specification of numbers.

However, using these interaction facilities is not always convenient: three-dimensional widgets allow more direct control on 3D transformations, while WYSIWYG-color editors may be more intuitive to select colors than selecting a color from a list of color names. On the other hand, it is important that the selection of a sensitive part in a caption leads to a uniform reaction. As a trade-off between dedicated interaction facilities and consistent feedback, the pop-up menus which are displayed after the selection can contain an item that invokes a dedicated editor, tailored to the interaction task.

The application of interactive figure captions reveals that it is not desirable to modify all parts of a figure caption. In particular, it is questionable whether modifications should be possible which would invalidate large portions of the caption. If the user, for instance, can alter the subject of the visualization, the rest of the caption becomes useless. As a rule of thumb, parts of the caption that describe *how* something is depicted should be interactive, while parts that describe *what* is depicted should not.

The interactive usage of captions is also limited by the fact that figure captions must correspond to generatable images. The system can only offer attributes for manipulation if it can actually handle an alternative specification for this attribute. If the system describes that one object is hidden by another one, the user cannot be allowed to delete this sentence from the figure caption or to replace the occluding object(s) , because it may be impossible to generate such a visualization.

4 Design and Realization of Figure Captions

In this section, we show how the content of figure captions[3] can be derived automatically based on the interaction with information spaces. Natural language generation in general comprises several stages: text planning, sentence planning, and sentence realization.

The main task of *text planning* is to decide which portion of the information space is to be expressed, considering parameters such as the intended user group, user preferences and the interaction history. It provides a representation of the structure of the text to be generated and an explicit or implicit description of the rhetorical structure, the intentional structure and the attentional structure [11].

Within the *sentence planning* step the text structure is broken down into sequences of sentences. This transformation must determine clause boundaries, syntactic structure of single clauses, theme of clauses, appropriate referential expressions (pronouns, definite or indefinite descriptions)

[3]The figure captions mentioned in the following refer to descriptive figure captions.

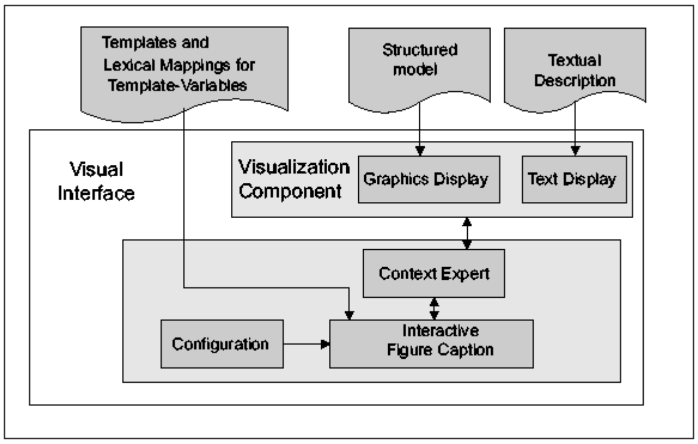

Figure 4: *Architecture of a visual interface with dynamic figure captions.*

within clauses, linguistic realizations (e.g., function word or phrases) that signal rhetorical relations, and lexicalization. Furthermore, the linear order of the sentences has to be determined.

The above mentioned specification from the sentence planning step is put into natural language expression within the *sentence realization*.

As all these tasks rely on the structure of the underlying information space, we will present the system's architecture and the internal data representation first. A detailed discussion of the techniques applied to generate figure captions automatically is given in an subsequent subsection. Finally, we present two figure captions realized as a consequence of an exemplary user interaction.

4.1 Data Representation and System Architecture

The information space consists of structured models and related textual information. The term *structured models* refers to geometric models that consist of distinct objects together with some semantic information, at least the names of objects and their affiliation to categories and sub-categories (e.g., musculus auricularis anterior, category: muscle, sub-category: eye muscle). Furthermore, cross-references between objects (e.g., muscles and bones which are located near to each other) are represented. Our system is based on geometric models which are commercially available. These have been manually enriched with information on object names and categories. A new approach to represent this domain knowledge in an external knowledge base is described in HARTMANN et al. ([13]).

The generation of figure captions that remain consistent with the image requires that all changes to a visualization are represented explicitly in data structures. Therefore, a special component is informed about changes in both *visualization components*, the *graphics display* as well as the *text display* and its initiator (user vs. system). This component manages the context of the visual interface and is therefore called the *context expert*. This terminology is related to the reference model for Intelligent Multimedia Presentation Systems [3]. The core of the reference model is an architectural scheme of the key components of multimedia presentation systems.

The context comprises the displayed labels and explanations, the viewing direction, colors, scaling factors of the displayed models and their objects together with the sequence of events, resulting in that visualization.

The *interactive figure caption* module analyzes the changes to the visualization and initiates the text generation based on the user's specification (the *configuration* in Figure 4). As discussed in Section 3.2 these user preferences influence which pieces of information are presented in the figure caption and when the generation of figure caption is triggered. As the visualization may be directly manipulated via interactive figure captions, this module has also to communicate with the context expert in order to hand over the requested changes to the visualization components.

For interactive 3D illustrations, one important aspect to comment concerns the visibility of objects after changing the viewing position. Though, the 3D model has to be analyzed as to what objects are hidden to what extent as well as which objects are now visible. Therefore, an off-line visibility analysis is employed: For a fixed number of viewing directions the visibility of parts of a complex object is analyzed. Assuming a constant distance between the camera position and the center of the model during user interaction, a fixed camera position is given for each viewing direction. For the given camera positions, rays are traced through the pixel of the rendered image. This method returns sequences of object hits which are used to estimate the relative visibility and the relative size of the projection for the parts of the model. Moreover, the list of occluding objects for a given part can be determined (e.g., $object_1$ is in front of $object_2$ and $object_3$ at *position (x,y)*). The relative visibility of a given part specifies the rate of rays, reaching it at first with respect to the rays crossing the object at all. Because this analysis is computationally expensive, a preprocessing of these values is employed for the predefined set of viewing directions, whereas the values of other viewing directions are estimated based on a linear interpolation between the recorded values. It turned out that for anatomic models visibility can be estimated well enough based on 26 predefined viewing directions which result from increasing azimuth- and declination angles in steps of 45 degrees.

4.2 Template-Based Generation

Although there is a high variation in the linguistic realization of the figure captions content, the basic structure is fairly fixed and some typical phrases dominate. Consequently, a template-based approach was used to generate figure captions within the ZOOM-ILLUSTRATOR. Templates are fixed natural language expressions which may contain variables. When a template is activated, the values of the template variables and an appropriate natural language expression describing them have to be determined.

Template-based generation methods require only a partial representation of the content to be conveyed. Furthermore, this method can be easily integrated into application programs as only little linguistic knowledge is required. On the other hand, large portions of text are fixed within templates. The texts produced with this method tend to be monotonous and the activation of the same templates with different values for template variables results in repetitive sentence structures. The large number of templates required for a flexible or multilingual realization of figure captions is a further disadvantage of the template-based generation method (Section 6 or [30] provides a detailed discussion of the advantages of template-based generation and linguistically motivated generation as well as their combinations).

[Please refer to page xx for this figure. To facilitate transmission, the large file for this figure is represented as a separate file]

Figure 5: *An image with complex abstraction and the figure caption describing them [27, p. 376].*

At first glance, it seems that the template-based generation is just a sentence-realization method. On the other hand, a system using a template-based approach has to decide which templates should be selected, to arrange them in a linear order and to estimate values for the template variables. We employ a *macro-structure* to select potentially important information which should be expressed within the figure caption. Furthermore, this macro-structure restricts the order while different templates within a *template category* can be used to flexibly describe the content of the structural elements within the macro-structure. Finally, within a *lexical mapping*, the linguistic realization for the template variables is determined. We will present some of these structures we derived from the analysis of figure caption in anatomy in turn.

To sum up, first of all the information space is restricted according to user preferences, i.e., only the most important changes are taken into consideration. A predefined macro-structure controls content determination, whereas the templates and the lexical mapping of template variables control the content realization.

Thus, the most important task in text planning, the content selection, is done via selection of the most important changes stored in the context expert and via conditions associated to structural elements in the macro-structure and the templates. Note that the macro-structure provides no explicit description of coherence relations between the structural elements. The selection of a template according to their constraining conditions is a simple kind of sentence planning, whereas the lexical mapping is responsible for the sentence realization.

4.2.1 Macro-Structure

We analyzed figure captions in various printed books in order to determine a *structure potential* of figure caption within a given domain. Within a macro-structure the structure of a particular text type is characterized in terms of typical parts (optional or obligatory *structural elements*) and the content that is conveyed by these parts. Furthermore, the macro-structure imposes restrictions on the order of structural elements.

As mentioned before, the content of a macro-structure depend on the domain as well as the text type. Thus, figure captions in textbooks of the application domain of the Zoom-Illustrator system, human anatomy, are carefully studied. In order to define a macro-structure which is general enough to be useful in other application domains, figure caption in scientific textbook are studied. Furthermore, we asked domain experts, i.e., undergraduates studying anatomy, what they expect to find in figure captions.

Anatomical atlases ([26] and [27]) consist of large, often complex images which are not surrounded by textual information, as it is the case in textbooks (hence the name "atlas"). Figure captions are the only form of textual information available and do not interfere with other references to an image.

Figure captions in anatomy follow a rather fixed structure. The first items mentioned are generally the name of the depicted contents, the viewing direction and the thematic focus (e.g., *muscles and ligaments of a foot from lateral*). This information is essential if an unfamiliar view of organs inside the human body is depicted. After the information about the image as a whole, important parameters of single objects are usually described. Among these parameters, manipulations that affect the visibility of objects are most important because they often influence the whole composition. If small objects are important in a specific context, they must be enlarged to emphasize them. In this case, the context is preserved for better orientation, so that the surrounding objects cannot be enlarged.

An example of a hand-drawn anatomic illustration and the accompanying figure caption is given in Figure 5. In the first sentence, the thematic focus of the illustration and the viewing direction with respect to the main object is provided. The second sentence informs the reader about radical changes due to invasion and removal of a number of objects.

Model View: the name of the depicted object and the viewing direction (4 templates),

Focus: the thematic focus of the illustration, information about different aspects of the illustration in case of multiple illustrations of the same object or symmetric object (5 templates),

Text View: the number of annotated objects of the thematic focus (2 templates),

Applied Illustration Technique: description of an abstraction in order to emphasize objects or to ensure visibility of objects from the thematic focus (4 templates),

Adaptive Zoom: description of selective enlargements of objects within the thematic focus (4 templates),

Occlusion: information about invisible objects within the thematic focus (4 templates),

Object Property: description of properties of objects with a high priority, as stated by the user (3 templates).

We represented this rather fixed structure in the macro-structure above. It starts with a description of the model view which includes the name of the model, the viewing direction and the aspects selected (the structural element *model view*). The structural element *text view* describes the filtering process of labels (whether all relevant labels could be displayed or not). This structural element was included to describe the effect of the zoom algorithm on labels (recall the discussion of the ZOOM-ILLUSTRATOR in Section 3.1).

Other structural elements (*applied illustration technique*, *adaptive zoom*, *occlusion* and *object property*) contain descriptions of the abstraction process, especially illustration techniques to emphasize objects, the description of rendering styles and attributes.

The relevance of *object properties* often depends on conventions in the domain. Presentation variables, for example, are often adapted to a specific context, e.g., to show spatial relations more clearly and to communicate which objects belong to the same category. In anatomy, objects of certain organ systems are colored uniformly according to accepted conventions. In figure captions, the use of color therefore needs to be described only if it differs significantly from the conventions.

As a result, these structural elements are arranged in a sequence which represents the order in which they are realized.

4.2.2 Template Categories

A template category characterizes the content of a structural element. It contains a collection of several conditional templates which covers frequently used phrases to convey the content of the structural element. Methods have to be supplied to estimate corresponding values for the template variables. These methods access information provided by the context expert and user preferences.

Templates selected to realize the content of the structural elements have to fulfill two conditions: their condition must be valid and for all template variables a valid substitution with values from the information space must be provided by the access methods. With this technique, several templates can be activated, or a template can be activated with different substitutions for the template variables. The structural elements *applied illustration technique*, *adaptive zoom*, *occlusion* and *object properties* is frequently realized by several applicable templates or multiple realizations of the same template. On the other hand, as the templates of the structural element *model view* have mutually excluding conditions, it is realized by exactly one template activation.

The ZOOM-ILLUSTRATOR can display two instances of a model (*Model*) with different viewing directions (*Direction*). Furthermore, different abstraction techniques can be applied to this model, resulting in different thematic foci or aspects (*AspectList*). Furthermore, visualizations may exploit the symmetry of anatomic objects, for example in a model of a human face, to show different aspects in both halves (see Figure 6).

Hence, different realizations are required, depending on whether one or two model views are presented simultaneously. When multiple aspects are displayed in several instances of a model or in a symmetric model, the similarities between the images are mentioned first, while the differences are described later. If horizontally arranged images are described by one caption, it is important that the left image be described before the right one, because there is a natural sequence of "reading images" from left to right. The templates to realize the structural element *model view* are presented below.

- The [*Model*] from [*Direction*].
- The [Model] from [$Direction_1$] and [$Direction_2$].
- The [*Model*] from [$Direction_1$] and [$Direction_2$] —on the left the
  [$AspectList_1$], on the right [$AspectList_2$].
- The [*Model*] from [*Direction*]—on the left the [$AspectList_1$], on the right [$AspectList_2$].

4.2.3 Lexical Mapping

The process of the sentence realization is reduced to a lexical mapping of the instances substituted for the template variables. When these instances represent numerical values, an interval is directly mapped to a sequence of words. With this approach, a certain value is always described in the same way. Colors are an example where this is useful. For this attribute, a widely accepted color naming system has been developed by Berk et al. [1] which can be used to describe colors objectively.

For the lexical realization of template variables, we need phrases to describe colors, transparency values and viewing directions. The naming of viewing directions considers conventions of the medical domain. In medical images, the frontal view, for example, is referred to as *ventral* which is more exact because ventral means *from the stomach.* Thus a reference point of the human body is used to name a viewing direction. By analogy, other viewing directions are named accordingly with reference to the human body.

In some cases, absolute mappings using fixed intervals are not suitable. Values are considered differently depending on the range of the values for this quantity in a given visualization; what is large in one context may be very small in another.

Furthermore, the lexical mapping may also exploit the information provided by the context expert. In order to refer to objects via their name, they must either be labeled or circumscribed by characteristic features which are easy to recognize in the image.

4.2.4 Dynamic and Interactive Figure Captions

The constraints of template categories and templates can be used to validate generated figure captions in dynamic figure captions in case of changes to the visualization (recall Section 3.2). If the conditions for a structural element or a template realizing the content of this structural element are violated, or the lexical mapping of a template variable results in another value, these structural elements or portions thereof are invalid.

Descriptions in natural language for the items of pop-up menus in interactive figure captions can be derived from the specification of the lexical mappings of the corresponding template variable. Using this approach, a modification of the external mapping specification is always kept consistent with the pop-up menu.

4.3 Example User Interaction and Figure Captions

We will illustrate the results from the above mentioned methods with two interactively created illustrations of the human face. In the first illustration (Figure 6) the user has manipulated the presentation so that the left and the right parts of the model look different. The figure caption indicates that the objects near the surface have been rendered semi-translucent in the right part of the model. Figure 7 shows another illustration of the same model. Here the user requested an explanation for a small muscle. The system has enlarged this muscle automatically to emphasize it. This side effect is reflected in the figure caption.

To realize the figure caption in Figure 6, the structural elements *model view* and *focus* were activated. The caption in Figure 7 was realized using the structural elements *model view*, *focus*, *text view* and *adaptive zoom*. In both captions, the structural element *model view* was realized with the first template from the template category.

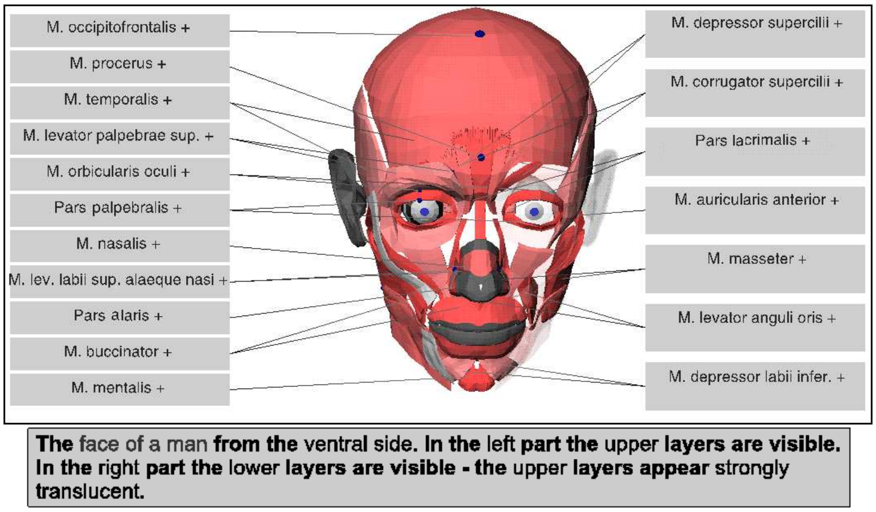

Figure 6: *An illustration conveying different thematic foci in both symmetric halves. Within the right half the objects of the upper layers where made semi-transparent to prevent occlusion of the lower layer objects. The application of this abstraction technique is described in the figure caption.*

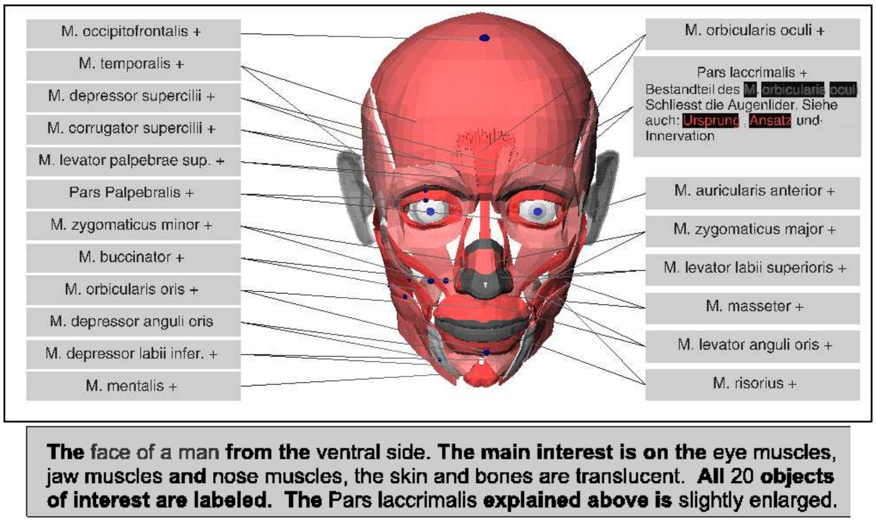

Figure 7: *The user's request for further information on a very small eye muscle via the selection of its label results in a system-initiated enlargement of that muscle. The first three sentences of the figure caption remain constant, whereas in the fourth sentence the effect of this abstraction technique is explained.*

5 Related Work

The incorporation of figure captions in interactive systems is a new approach to enhance the usability of visual interfaces. Figure captions in a narrow sense—captions that describe images previously generated—were first developed by MITTAL et al. [18]. This work has been carried out in the context of the SAGE project, where complex diagrams are generated. MITTAL et al. argue that users can deal with more complex diagrams if they are accompanied by explanatory captions. In other words: complex relations that must be depicted in several images (without captions) can be integrated in one diagram when explained appropriately.

Some systems produce what we introduced as instructive figure captions (recall Section 2.2). In particular, in the WIP project (Knowledge-Based Information Presentation) [36] and in the COMET project (CO-ordinated Multimedia Explanation Testbed) [9] technical illustrations and textual descriptions are generated to explain the repair and maintenance of technical devices. The content of their accompanying texts often goes beyond the content of the figure. The textual components are generated based on large knowledge bases, describing object properties and actions. To realize instructive figure captions, an explicit representation of actions is required. As it is time-consuming to create those representations, practical realizations have a restricted application domain.

The selected content is presented within different media (graphics, animation and text) which are used either complementary or redundantly. The redundant information presentation aims at reinforcing the message, e.g., by appropriate cross-references. In this *media selection* process, the suitability of a medium to convey information is considered. Moreover, sophisticated *media coordination* facilities are employed.

The approach of interactive figure captions differs from the media coordination in WIP and COMET. In our concept, either the caption describes an image *or* the image is guided by the caption. The generation processes in the ZOOM-ILLUSTRATOR are still autonomous, whereas graphics and text generation mutually influence each other in WIP and COMET. In contrast to these systems, we target the interactive exploration of visualizations and not a final presentation.

Furthermore, in these systems changes which a user may make in an image are not reflected in the figure caption, since these application address issues not related to user interactions.

6 Discussion

We now turn to a discussion of some of the limitations and problems associated with describing abstraction.

Figure captions vs. animation to communicate abstraction. To facilitate the recognition of abstraction techniques, animation can be employed to interpolate between the original values and the values resulting from the application of abstraction techniques. With such animation the user can watch what happens instead of being forced to interpret a radical new image. Furthermore, with these animations the user's attention is directed to the distortion introduced by the abstraction.

Figure captions are more important for static images resulting from an abstraction, but they are also useful to summarize an abstraction in different levels of detail and focused on different aspects. Furthermore, causalities cannot be expressed by animations (e.g., objects are displayed transparently

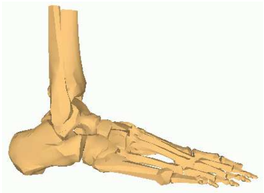

Figure 8: *Bones of the foot, lateral view.*

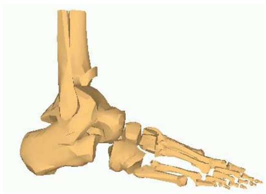

Figure 9: *Bones of the foot, lateral view; calcaneus strongly enlarged.*

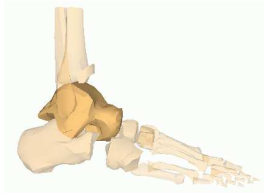

Figure 10: *Bones of the foot, lateral view; calcaneus strongly enlarged. The transparency encodes the scaling factor.*

Figure 11: *Visualization of the distortion introduced by abstraction and descriptive figure captions. The left illustration shows the bones of a human foot in original scale. The middle illustration is focused on a bone in the center using 3D Zoom which causes an enlargement of this bone, while all other objects shrink in order to fit on the same space. In the right illustration the scaling factor is mapped onto transparency, (by kind permission of Andreas Raab).*

as they occlude focused objects). The combination of animation and figure captions is currently being investigated and promises to reduce the complexity of descriptive figure captions. Nevertheless, figure captions remain an efficient means to direct the user's attention to important aspects in the animation. The generation of figure captions for animation is an interesting point for future work.

Visualization of the distortion introduced by abstraction. The analysis so far shows the importance of describing the distortion introduced by *some* abstraction techniques. Additional presentation parameters like transparency may be used to visualize the geometric distortion introduced by abstraction methods in the rendered image. Grey-scale values are employed in Carpendale et al. [5], for example, to communicate the extent to which a fish-eye view differs from the corresponding normal view. As this is just another abstraction, its effect could be described in figure captions.

As mentioned in Section 2.1, the selective scaling of objects may lead to a misinterpretation of the resulting image. The 3D Zoom (recall discussion in Section 3.1) is a very powerful abstraction technique which introduces a scaling of *all* objects within the illustration. In the middle part of Figure 11, one bone is emphasized using 3D Zoom, whereas the left part of that illustration shows an unscaled rendition of the model. To emphasize the distortion introduced by the application of the 3D Zoom, the scaling is mapped on transparency values in the right illustration of Figure 11.[4]

Indexing of rendered images with figure captions. As a figure cap-

[4]The images within this section are furnished with hand-made figure captions following the macro-structure presented in Section 4.2.1.

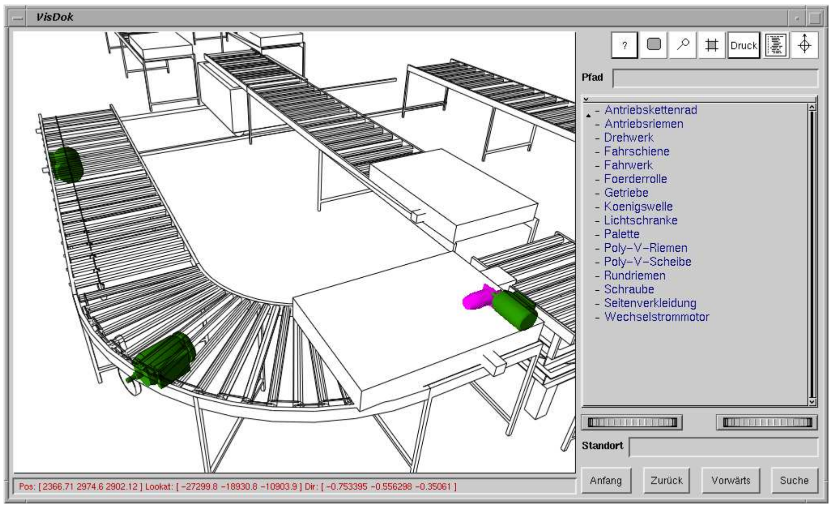

Figure 12: *The coding area of a package distribute center with focus on motors.*

Figure 13: *An illustration of a complex technical device and a descriptive figure caption in technical documentation.*

tion summarizes the content of an illustration, they can be used as an index within multimodal information retrieval [33]. Thus, the automatic generation of a figure caption for screen-shots of interactively created computer graphics may serve as an automatically created index. As these illustrations are interpreted in the absence of the interactive situation in which they have been generated, they should summarize the interaction applied.

Descriptive figure captions in technical documentation. An interesting area for future work is the application of the ideas developed in this paper to technical documentation. In this area, 3D models resulting from the actual construction process are used to support people who repair and maintain technical devices. We developed a system for the interactive exploration of technical documentation ([12] and [14]) which uses a number of abstraction techniques to emphasize objects. In Figure 13 a combination of line-drawing and shading is used in order to focus the illustration on the motors. As again 3D models are involved, similarities between this area and medical illustrations exist.

Empirical evaluation of abstraction techniques. The reflection of abstraction techniques in figure captions provide some hints on their importance. Further work must incorporate an empirical evaluation in order to derive a fine-grained evaluation of abstraction techniques.

Furthermore, empirical evaluation should analyze the importance of potential structural elements of the macro-structure (recall Section 4.2) in figure caption for different user groups and parameters to which the visual interface may be adopted to.

Recently, a first evaluation of the interaction techniques used by the Zoom-Illustrator for the explanation of spatial phenomena was carried out ([21]). One goal was to evaluate whether undergraduates studying medicine find figure captions useful for a correct interpretation of anatomic illustrations generated interactively by the Zoom-Illustrator. On a scale between 0 (redundant) and 10 (most indispensable) all 9 subjects ranked the usefulness between 9 and 10 (9.8 on average). No other feature of the Zoom-Illustrator reached those ratings. Despite the small number of subjects, that is a very impressive result which clearly emphasizes the importance of figure captions, also in interactive environments.

Refinements of the realization of figure captions. In the current implementation, we employ the template-based approach for the realization of figure captions. As we pointed out, this may lead to monotonous text. To overcome this, the number of different templates within a template category could be enlarged. When figure captions in different target languages are required, the number of templates increases further, resulting in high maintenance efforts ([30]).

Linguistically motivated generation methods are appropriate for the generation of flexible text, as these methods concentrate on the generation of coherent and cohesive texts using a large variety of syntactic structures. To establish cohesion in text appropriately, choices of definite descriptions (pronouns, nominal groups) and the aggregation of parallel structures are an important means. However, for these methods a formal representation (domain knowledge, user models and a representation of the context) and large linguistic resources (grammars and lexica) are necessary, which leads to a high initial overhead.

To increase the flexibility of the generated texts, methods from linguistically motivated generation should be combined with the template-based approach (see [38] and [4]). There are different representation formalisms for describing the text structure, like static schema ([17]) or the functional-oriented rhetorical structure theory ([16]). Some text planning techniques (see for example [19]) create an explicit representation of the text structure which cover the rhetorical, intentional and in a few cases also the attentional structure of the text to be generated. A text planning technique is already employed in the Visdok project [12]. In this enhanced framework, text operators will replace the structural elements and their specification. The text structure can be utilized to determine how different templates can be combined in order to generate a more flexible text. Furthermore, the text structure could be used for the lexical realization of template variables, (see [6], [29], and [15]) and the aggregation of parallel structures.

Text-to-speech. In a learning scenario with permanent user interaction, the user will focus on the graphics and thus might not notice the content of figure captions or even dynamic changes in them. In Section 3.2 incremental additions to existing figure captions were proposed to facilitate the recognition of new parts of the figure caption. But this method does not overcome the conflict of competing content in different output media. Spoken comments, however, can be perceived by the user in parallel. This would require a modification of the realization as to the peculiarities of spoken output. The fixed structure on which both generation methods are based would lead to the constant repetition of large parts of the comments. Furthermore, the problem resulting from redundant information (as mentioned in the last paragraph) would increase. Hence, the spoken comments should mention only system-initiated changes in response to user interaction, while the full content of the figure caption should only be conveyed at user's request.

7 Summary

In this paper, we presented an approach to furnish computer generated illustrations with automatically generated, interactive figure captions. These figure captions describe the depicted content, the view on an underlying model and the thematic aspects on which the visualization focuses. Several abstraction techniques for graphical emphasis of focused objects are used interactively by the ZOOM-ILLUSTRATOR and described in figure captions. The interactive exploration of anatomic models is supported by dynamic, adaptable and interactive figure captions.

Although inspired mainly by anatomic illustrations, the concepts presented here are broadly applicable to describe visualizations of structured data, as for example maps [22] and technical documentation [14].

Acknowledgments

The authors would like to thank those persons who contributed with their concepts and implementations to the work presented in this paper: Torsten Sommerfeld (figure captions for anatomic illustrations), Alf Ritter, Michael Rüger and Andreas Raab (zoom techniques in 2D and 3D). Moreover, we wish to thank Rainer Michel for fruitful discussions; he also contributed figure captions for computer generated maps for blind people. We wish to thank Brigitte Grote and Sylvia Zabel for a thorough reading and many suggestions to improve the paper. Finally, we would like to thank Ehud Reiter and three anonymous reviewers for their comments and suggestions.

References

[1] T. Berk, L. Brownstone, and A. Kaufmann. A New Color-Naming System for Graphics Languages. *IEEE Graphics and Applications*, 2(3):37–44, May 1982.

[2] R. M. Bernard. Using Extended Captions to Improve Learning from Instructional Illustrations. *British Journal of Educational Technology*, 21(3):215–225, 1990.

[3] M. Bordegoni, G. Faconti, M. T. Maybury, T. Rist, S. Ruggieri, P. Trahanias, and M. Wilson. [A Reference Model for Intelligent Multimedia Presentation Systems][5]. *Computer Standards & Interfaces: The International Journal on the Development and Application of Standards for Computers, Data Communications and Interfaces*, 6,7(18):477–496, Dec. 1997.

[4] S. Busemann. [Best-First Surface Realization][6]. In D. Scott, editor, *Proc. of the 8th International Workshop on Natural Language Generation*, Herstmonceux, Sussex, Great Britain, June, 13–15 1996.

[5] M. S. T. Carpendale, D. J. Cowperthwaite, and F. D. Fracchia. 3-Dimensional Pliable Surfaces: For the Effective Presentation of Visual Information. In *8th Annual Symposium on User Interface Software and Technology (UIST'95)*, pages 217–226, Pittsburg, Pennsylvania, Nov., 14–17 1995. ACM Press, New York.

[5] Hyperlink: http://www.dfki.de/ rist/csi97/SRM.ps.gz
[6] Hyperlink: http://xxx.lanl.gov/abs/cmp-lg/9605010

[6] R. Dale and E. Reiter. [Computational Interpretations of the Gricean Maxims in the Generation of Referring Expressions][7]. *Cognitive Science*, 19(2):233–263, 1995.

[7] J. Dill, L. Bartram, A. Ho, and F. Henigman. A Continuously Variable Zoom for Navigating Large Hierarchical Networks. In *Proc. of IEEE Conference on Systems, Man and Cybernetics*, pages 386–390, 1994.

[8] S. K. Feiner. APEX: An Experiment in the Automated Creation of Pictoral Explanations. *IEEE Computer Graphics and Applications*, 4(11):29–38, Nov. 1985.

[9] S. K. Feiner and K. R. McKeown. Automating the Generation of Coordinated Multimedia Explanations. In M. T. Maybury, editor, *Intelligent Multimedia Interfaces*, pages 117–138. AAAI Press, Menlo Park, CA, 1993.

[10] E. H. Gombrich. *The Image and the Eye: Further Studies in the Psychology of Pictorial Representation.* Cornell University Press, Ithaca, NY, 1982.

[11] B. J. Grosz and C. L. Sidner. Attention, Intentions and the Structure of Discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[12] K. Hartmann, R. Helbing, D. Rösner, and T. Strothotte. [Visdok: Ein Ansatz zur interaktiven Nutzung von technischer Dokumentation][8]. In P. Lorenz and B. Preim, editors, *Simulation und Visualisierung '98*, pages 308–321, Magdeburg, Germany, Mar., 5–6 1998. SCS–Society for Computer Simulation Int., Erlangen.

[13] K. Hartmann, A. Krüger, S. Schlechtweg, and R. Helbing. [Interaction and Focus: Towards a Coherent Degree of Detail in Graphics, Caption and Text][9]. In O. Deussen, V. Hinz, and P. Lorenz, editors, *Simulation und Visualisierung '99*, pages 127–137, Magdeburg, Germany, Mar., 4–5 1999. SCS–Society for Computer Simulation Int., Erlangen.

[14] R. Helbing, K. Hartmann, and T. Strothotte. [Using Dynamic Visual Emphasis in Interactive Technical Documentation][10]. In *Proc. of ECAI'98 Workshop on Combining AI and Graphics for the Interface of the Future*, Brighton, UK, Aug., 24 1998.

[15] H. Horacek. A New Algorithm for Generating Referential Descriptions. In W. Wahlster, editor, *Proc. of the 12th European Conference on Artificial Intelligence (ECAI'96)*, pages 577–581, Budapest, Hungary, Aug., 11–16 1996. John Wiley & Sons, Chichester.

[16] W. C. Mann and S. A. Thompson. Rhetorical Structure Theory: A Theory of Text Organization. In L. Polanyi, editor, *The Structure of Discourse*. Ablex, Norwood, N.J., 1987.

[17] K. R. McKeown. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text.* Cambridge University Press, Cambridge, 1985.

[7] Hyperlink: http://xxx.lanl.gov/abs/cmp-lg/9504020

[8] Hyperlink: http://www-ai.cs.uni-magdeburg.de/ hartmann/Papers/sv98.ps.gz

[9] Hyperlink: http://www-ai.cs.uni-magdeburg.de/ hartmann/Papers/sv99.ps.gz

[10] Hyperlink: http://www-ai.cs.uni-magdeburg.de/ hartmann/Papers/visdok-ecai98.ps.gz

[18] V. O. Mittal, J. D. Moore, G. Carenini, and S. Roth. Describing Complex Charts in Natural Language: A Caption Generation System. *Computational Linguistics*, 24(3):431–467, 1998.

[19] J. D. Moore and C. L. Paris. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics*, 19(4):651–694, 1993.

[20] E. G. Noik. A Space of Presentation Emphasis Techniques for Visualizing Graphs. In W. A. Davis and B. Joe, editors, *Proc. of Graphics Interface'94*, pages 225–233, Banff, Alberta, Canada, May, 18–20 1994. Canadian Human-Computer Communications Society.

[21] I. Pitt, B. Preim, and S. Schlechtweg. Evaluating Interaction Techniques for the Explanation of Spatial Phenomena. In U. Arend, E. Eberleh, and K. Pitschke, editors, *Software-Ergonomie '99 — Design von Informationswelten*, pages 275–286, Walldorf, Germany, Mar., 8–11 1999. B. G.Teubner Stuttgart.

[22] B. Preim, R. Michel, K. Hartmann, and T. Strothotte. [Figure Captions in Visual Interfaces][11]. In T. Catarci, M. F. Costabile, S. Levialdi, and L. Tarantino, editors, *Advanced Visual Interfaces: An International Workshop, AVI'98*, pages 235–246, LÁquila, Italy, May, 24–27 1998. ACM Press, New York.

[23] B. Preim, A. Raab, and T. Strothotte. [Coherent Zooming of Illustrations with 3D-Graphics and Text][12]. In W. E. Davis, M. Mantei, and V. Klassen, editors, *Proc. of Graphics Interface '97*, pages 105–113, Kelowna, BC, Canada, May, 19–23 1997. Canadian Human-Computer Communications Society.

[24] B. Preim, A. Ritter, and T. Strothotte. [Illustrating Anatomic Models — A Semi-Interactive Approach][13]. In K.-H. Höhne and R. Kikinis, editors, *Visualization in Biomedical Computing: 4th International Conference (VCB'96)*, pages 23–32, Hamburg, Germany, Sept., 23–25 1996. Springer Verlag, Berlin.

[25] B. Preim, A. Ritter, T. Strothotte, D. R. Forsey, and L. Bartram. [Consistency of Rendered Images and Their Textual Label][14]. In H. P. Santo, editor, *Proc. of CompuGraphics' 95*, pages 201–210, Alvor, Portugal, Dec., 6–10 1995.

[26] R. Putz and R. Pabst, editors. *Sobotta; Atlas der Anatomie des Menschen.* Urban & Schwarzenberg, München, Wien, Baltimore, 1993. 20. Auflage.

[27] R. Putz and R. Pabst, editors. *Sobotta; Atlas of Human Anatomy*, volume 2, Thorax, Abdomen, Pelvis, Lower Limb. Williams & Wilkins, Baltimore, 1997. 12th English Edition.

[28] A. Raab and M. Rüger. [3D-Zoom: Interactive Visualisation of Structures and Relations in Complex Graphics][15]. In B. Girod, H. Niemann,

[11] Hyperlink: http://www.cevis.uni-bremen.de/ bernhard/papers/avi98.ps.gz

[12] Hyperlink: http://isgwww.cs.uni-magdeburg.de/ bernhard/papers/gi97.html

[13] Hyperlink: http://isgwww.cs.uni-magdeburg.de/ bernhard/papers/hamburg.html

[14] Hyperlink: http://isgwww.cs.uni-magdeburg.de/ bernhard/papers/compualf.html

[15] Hyperlink: http://isgwww.cs.uni-magdeburg.de/ raab/papers/erlangen/

and H.-P. Seidel, editors, *3D Image Analysis and Synthesis*, pages 125–132, Erlangen, Germany, Nov., 17–18 1996. infix, Sankt Augustin.

[29] E. Reiter, C. Mellish, and J. Levine. [Automatic Generation of Technical Documentation][16]. *Applied Artificial Intelligence*, 9:259–287, 1995.

[30] E. Reiter and E. Mellish. [Optimizing the Costs and Benefits of Natural Language Generation][17]. In *Proc. of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 1164–1169, Chambery, France, 1993. Morgan Kaufmann.

[31] T. Rist, T. Krüger, G. Schneider, and D. Zimmermann. AWI – A Workbench for Semi-Automated Illustration Design. In *Proc. of Advanced Visual Interfaces (AVI '94)*, pages 59–68., Bari, Italy, 1994. ACM Press, New York.

[32] D. D. Seligmann and S. K. Feiner. Automated Generation of Intent-Based 3D Illustrations. *Computer Graphics*, 25(4):123–132, 1991.

[33] R. K. Srihari and D. T. Burhans. Visual Semantics: Extracting Visual Information From Text Accompanying Pictures. In *Proc. of Annual National Conference on Artificial Intelligence (AAAI'94)*, pages 793–798, Seattle, Washington, Aug., 1–4 1994. AAAI Press, Menlo Park, CA.

[34] C. Strothotte and T. Strothotte. *Seeing Between the Pixels: Pictures in Human-Computer Interaction.* Springer Verlag, Berlin, 1997.

[35] T. Strothotte et al. *Computational Visualization: Graphics, Abstraction, and Interactivity.* Springer Verlag, Berlin, 1998.

[36] W. Wahlster, E. André, W. Finkler, H.-J. Profitlich, and T. Rist. Plan-Based Integration of Natural Language and Graphics Generation. *Artificial Intelligence*, 63:387–427, 1993.

[37] B. Weidenmann. Informative Bilder (Was sie können, wie man sie didaktisch nutzt und wie man sie nicht verwenden sollte). *Pädagogik*, 1989(9):30–34, September 1989.

[38] M. White and T. Caldwell. [EXEMPLARS: A Practical, Extensible Framework for Dynamic Text Generation][18]. In E. Hovy, editor, *Proc. of the 9th International Workshop on Natural Language Generation*, pages 266–275, Niagara-on-the-Lake, Canada, Aug., 5–7 1998.

[16] Hyperlink: http://xxx.lanl.gov/abs/cmp-lg/9411031

[17] Hyperlink: http://www.dai.ed.ac.uk/staff/personal_pages/chrism/papers/ijcai-93-nlg.ps

[18] Hyperlink: http://www.cogentex.com/papers/exemplars-inlgw98.pdf