

# Bildunterschriften zur Erklärung von räumlichen Zusammenhängen

*Knut Hartmann    Bernhard Preim    Torsten Sommerfeld    Thomas Strothotte*

*Otto-von-Guericke-Universität Magdeburg  
Fakultät für Informatik  
Postfach 41 20, D-39016 Magdeburg*

*e-mail: hartmann@iik.cs.uni-magdeburg.de  
{bernhard|sommerfe|tstr}@isg.cs.uni-magdeburg.de*

## 1 Motivation

Eines der Ziele multimedialer Präsentationssysteme ist es, computergenerierte sprachliche und bildliche Ausgaben geeignet zu kombinieren. Dabei sollte eine dynamische Aufteilung der Informationen auf die verschiedenen Präsentationskanäle in Abhängigkeit von den zu vermittelnden Sachinhalten erfolgen, um eine effektive Interaktion mit Endbenutzern zu ermöglichen. Anwendungen hierfür liegen beispielsweise im Bereich von Lernsoftware für Medizinstudenten.

Für die Vermittlung komplexer räumlicher Zusammenhänge sind Bilder von großem Vorteil, da ein reichhaltiges Repertoire an Mitteln zur wissenschaftlichen Illustration zur Verfügung steht (z.B. Liniengraphiken, Abstraktion, Transparenz, Wegklappen von verdeckenden Teilen; siehe [Strothotte und Wagener, 1997]). Gerade wegen dieser Vielfalt an graphischen Techniken können aber rein bildliche Darstellungen komplexer räumlicher Zusammenhänge wiederum schwer verständlich sein, da der Betrachter nicht immer von der Darstellung auf die Eigenschaften des Dargestellten schließen kann. Wird beispielsweise ein Teilobjekt vergrößert dargestellt, damit seine Details besser zu erkennen sind, so weiß der Betrachter nicht unbedingt, wie groß dieses Teilobjekt in der Realität ist.

Das Phänomen, daß „kein Bild sich selbst erklärt“ [Gombrich, 1984, S.142] wurde in der Literatur bereits ausführlich diskutiert (siehe z.B. [Weidenmann, 1989]). Es ist unstrittig, daß Bilder durch sprachnahe Informationen begleitet werden müssen. Allerdings ist dieses Problem im Zusammenhang mit 3D-Graphiken derart komplex, daß wir eine Lösung nur mittelfristig anstreben können. Ein in seiner Komplexität reduzierteres, aber ähnliches Vorläuferproblem ist es, textuelle Bildunterschriften zu dynamisch generierten 3D-Graphiken automatisch zu erzeugen. Dieses reduzierte Problem hat den Vorteil, daß die Techniken des wissenschaftlichen Illustrierens, wie sie in Lehrbüchern zum Einsatz kommen, maschinell beherrschbar geworden sind und somit die Bildunterschriften aus diesen Lehrbüchern als Vorbilder herangezogen werden können.

In diesem Artikel wird die Frage der Generierung von textuellen Bildunterschriften zur Begleitung von komplexen, dynamisch erzeugten 3D-Graphiken thematisiert. In Kapitel 2 wird die Struktur und der Inhalt von Bildunterschriften näher untersucht. Im Kapitel 3 wird anhand eines implementierten Prototyps auf die zugrundeliegenden Repräsentationen und die template-basierte Generierung von Bildunterschriften eingegangen.

Auf Besonderheiten, die sich aus dem Einsatz von Bildunterschriften in einer interaktiven Umgebung ergeben, wird in Kapitel 4 eingegangen. In den Kapiteln 5 und 6 werden die Ergebnisse diskutiert und zusammengefaßt.

## 2 Struktur und Inhalt von Bildunterschriften

Bilder in herkömmlichen Medien umfassen neben der Darstellung von Objekten auch Annotationen von Teilobjekten, die ihre sprachliche Bezeichnung enthalten, sowie die Bildunterschrift. In diesem Kapitel gehen wir darauf ein, wie der Inhalt der Bildunterschrift von der gezeigten Illustration *und* seinen Annotationen abhängig ist.

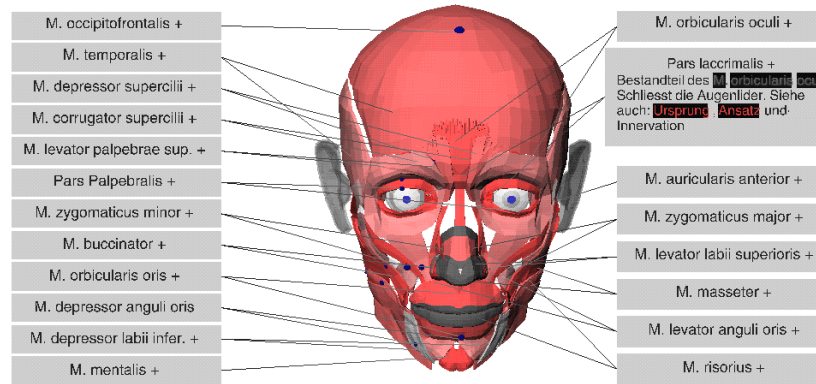
Bildunterschriften wurden in pädagogischen und psychologischen Untersuchungen klassifiziert. So unterscheiden wir in Anlehnung an Bernard [Bernard, 1990] zwischen deskriptiven und instruktiven Bildunterschriften. *Deskriptive Bildunterschriften* halten wichtige Aussagen eines Bildes in verständlicher sprachlicher Form fest, während in *instruktiven Bildunterschriften* vermittelt wird, wie bestimmte Handlungen auszuführen sind.

Bei der Spezifikation des Inhaltes und der Struktur orientieren wir uns an deskriptiven Bildunterschriften, wie sie in Lehrbuchttexten der Technik und Anatomie vorkommen. Besonders interessant sind Bildunterschriften in anatomischen Atlanten (beispielsweise [Staubesand, 1988]). Die Bilder sind dort nicht in einen umgebenden Fließtext eingebettet, so daß die Bildunterschriften besonders ausführlich sind.

In der anatomischen Domäne werden oftmals zur besseren Darstellung fokussierter Teile andere Objekte entfernt, durchtrennt oder weggeklappt. Weiterhin finden wir Schnitte durch ein Objekt. In der technischen Domäne werden häufig Explosionsdarstellungen oder schematische Darstellungen verwendet. Häufig wird auch der Zweck dieser Veränderungen erläutert (z.B.  $\langle \text{Objekt}_1 \rangle$  entfernt, um  $\langle \text{Objekt}_2 \rangle$  zu zeigen). Obwohl sich unsere Beobachtungen auf anatomische Atlanten konzentrieren, lassen sie sich teilweise auf die Erklärung räumlicher Zusammenhänge in anderen Bereichen verallgemeinern, da auch dort der Einsatz illustrativer Mittel zu komplizierten Bildern führen kann, die in der Bildunterschrift reflektiert werden.

Die Struktur der in anatomischen Bildatlanten verwendeten Bildunterschriften ist relativ starr. Sie enthalten zumindestens einen Bildidentifikator und den Bildbezeichner. Der *Bildidentifikator* ist ein eindeutiger Bezeichner, der zum referentiellen Gebrauch im umgebenden Fließtext bzw. in folgenden Bildunterschriften sowie zur Katalogisierung dient. Im *Bildbezeichner* wird das dargestellte Objekt benannt und die Sichtrichtung aufgeführt (z.B. *Fuß des Menschen, Ansicht von vorn*).

Objekte können nach bestimmten Kriterien *klassifiziert* werden. In der Anatomie werden Objekte nach der Zugehörigkeit zu Organsystemen (relevant in unserem Beispiel sind Knochen, Muskeln und Bänder) klassifiziert. Die Bildunterschrift kann darauf hinweisen, welche Objektklassen in der Illustration durch Annotationen oder die Anwendung illustrativer Mittel hervorgehoben werden. Weitere Teile der Bildunterschrift kommentieren den Einsatz gestalterischer Mittel. Gezielte Veränderungen einzelner Objekte werden in Bildunterschriften erwähnt, um das mit diesen Veränderungen verfolgte Ziel anzuzeigen.



**The face of a man from the ventral side. The main interest is on the eye muscles, jaw muscles and nose muscles, the skin and bones are translucent. All 20 objects of interest are labeled. The Pars lacrimalis, which is colored deep red, is solid. Pars lacrimalis is slightly enlarged.**

Abbildung 1: *Bildschirmansicht des ZOOM ILLUSTRATORS*

Der Einsatz von Farben ist stark konventionalisiert. Diese Konventionen werden — wenn überhaupt — einmalig erläutert. Ein Abweichen von der Konvention, um ein Objekt besonders hervorzuheben oder einen bestimmten Kontrast sicherzustellen, bedarf deshalb einer Erläuterung in der Bildunterschrift.

In der Bildunterschrift kann nur auf Objekte verwiesen werden, die in der bildlichen Darstellung sichtbar sind *und* deren Bezeichnung durch Annotationen angegeben ist. Diese Regel wird allerdings nicht angewendet, wenn durch den Einsatz gestalterischer Mittel Objekte nicht in der Illustration dargestellt werden, obwohl sie aus dem gegebenen Blickwinkel sichtbar sind.

Die Orientierung an Bildunterschriften in statischen Lehrmedien hat aber ihre Grenzen. In interaktiven Systemen können Bildunterschriften auf den konkreten Benutzer zugeschnitten werden. Interaktive Veränderungen der Sichtrichtung des dargestellten Modells oder seiner Darstellungsweise können aber dazu führen, daß Bildunterschriften unvollständig werden oder gar den im Bild dargestellten Sachverhalt nicht widersprechen, also die Konsistenz zwischen einem Bild und seiner Unterschrift nicht mehr gewahrt ist. Auf diese Aspekte von Bildunterschriften in interaktiven Medien wird in Kapitel 4 näher eingegangen.

### 3 Template-basierte Realisierung von Bildunterschriften

Bildunterschriften wurden in ein interaktives Illustrationssystem, den ZOOM ILLUSTRATOR [Preim et al., 1996], integriert (siehe auch [Sommerfeld, 1997]). In diesem System kann eine beliebige Sichtrichtung auf das dargestellte 3D-Modell gewählt und die Präsentation auf eine Menge besonders fokussierter Objekte zugeschnitten werden. Auf Benutzerinteraktionen reagiert das System, indem der Bild-Text-Zusammenhang durch eine aktualisierte Bildunterschrift aufrechterhalten wird. Die Initiative zur Erzeugung der aktualisierten Bildunterschrift kann dabei vom Benutzer oder vom System selbst ausgehen.

Zur Generierung der Bildunterschrift wenden wir aufgrund der relativ festen Struktur der Bildunterschriften ein template-basiertes Verfahren an. Die Templates sind in Gruppen zusammengefaßt, die zur Generierung der einzelnen Abschnitte der Bildunterschrift (des Bildbezeichners, der fokussierten Objekte oder Objektklassen, der Kommentierung des Einsatzes gestalterischer Mittel) dienen. Unser Prototyp geht von 15 Templates aus, die kontextbezogen ausgewählt werden und mit aktuellen Werten für die in den Templates enthaltenen Variablen gefüllt werden. Abbildung 1 zeigt eine durch den ZOOM ILLUSTRATOR interaktiv erstellte Illustration.

Das folgende Template wird verwendet, um den ersten Satz der Bildunterschrift — den Bildbezeichner — in Abbildung 1 zu generieren.

Object+Direction  
The [Object] from the [Direction].

Die Variablen **Object** und **Direction** werden mit der sprachlichen Bezeichnung des dargestellten Objektes und einem sprachlichen Ausdruck für die aktuelle Sichtrichtung belegt. Im ZOOM ILLUSTRATOR wird die Kamera stets auf das Zentrum des Modells gerichtet, so daß die Sichtrichtung durch einen Azimut- und einen Deklinationswinkel repräsentiert werden kann (siehe Abbildung 2). Der Azimutwinkel gibt den Längengrad ( $-180^\circ$  bis  $180^\circ$ ), die Deklination den Breitengrad ( $-90^\circ$  bis  $90^\circ$  mit  $0^\circ$  als Äquator) an.

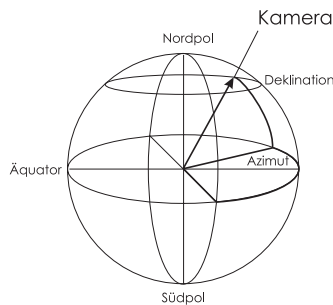


Abbildung 2: Einheitskugel zur Bestimmung der Sichtrichtung

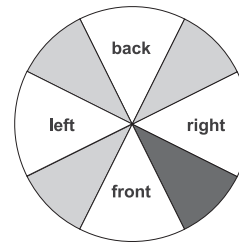


Abbildung 3: Zuordnung der Sichtrichtung – Draufsicht auf den Äquator

Die interne Repräsentation der Sichtrichtung wird anhand der in Abbildung 4 angegebenen Transformation in eine sprachliche Bezeichnung überführt. Zu beachten ist die Überschneidung der Zuordnung von Azimut- und Deklinationswerten. Abbildung 3 verdeutlicht diese Überschneidungen für den Azimutwert. In den dunklen Zonen werden dem

geometrische Repräsentation		sprachliche Umsetzung
Deklination	$[45^\circ, 90^\circ]$	top
Deklination	$[-45^\circ, -90^\circ]$	bottom
Azimut	$[-50^\circ, 50^\circ]$	front
Azimut	$[-130^\circ, -180^\circ]$ oder $[130^\circ, 180^\circ]$	back
Azimut	$[-40^\circ, -140^\circ]$	left
Azimut	$[40^\circ, 140^\circ]$	right

Abbildung 4: Sprachliche Realisierung von Blickwinkeln

<object <sub>1</sub> > is larger than <object <sub>2</sub> >.	Relative Größe von <object <sub>1</sub> > und <object <sub>2</sub> >
<object <sub>1</sub> > is partly covered by <object <sub>2</sub> >. <object <sub>1</sub> > is not visible. <object <sub>1</sub> > is not visible, because it is situated behind <object <sub>2</sub> >.	Verdeckungsinformationen
The <object <sub>1</sub> > is placed in the <position> of the model.	Relative Position von <object <sub>1</sub> > bezüglich der aktuellen Sichtrichtung.

Abbildung 5: Ausnutzung der Sichtbarkeitsinformationen in der Bildunterschrift

Azimutwert zwei sprachliche Realisierungen zugeordnet. Beispielsweise wird der dunkel hervorgehobene Bereich mit ‘front right side’ bezeichnet. Analog wird die Bezeichnung ‘top’ oder ‘bottom’ in Abhängigkeit vom aktuellen Deklinationswert hinzugefügt. In der Anatomie werden die Sichtrichtungen mit speziellen Termini bezeichnet, in Abbildung 1 wird die Sichtrichtung deshalb statt mit ‘front side’ als *ventral* bezeichnet.

Der zweite Abschnitt der Bildunterschrift in Abbildung 1 gibt an, auf welchen Objekten oder Objektklassen das Hauptinteresse des Betrachters liegt. Diese können durch den Benutzer ausgewählt werden. Die Templates, die zur Generierung dieses Abschnitts der Bildunterschrift verwendet werden, greifen ebenfalls auf diese Informationen zu. Weiterhin wird in der Bildunterschrift angegeben, ob alle Objekte der fokussierten Objektklassen in der Darstellung mit Annotationen versehen werden konnten.

Die untersuchten Bildunterschriften in anatomischen Atlanten enthalten keinen Bezug darauf, ob alle oder nur ein Teil der fokussierten Objekte mit Annotationen versehen wurden. Diese Information in der Bildunterschrift wurde in unseren Prototypen aufgenommen, weil aufgrund von Platzrestriktionen nicht immer alle fokussierten Objekte annotiert werden können, obwohl durch Anwendung von Fisheye-Techniken [Preim et al., 1997] der verfügbare Platz für Annotationen sehr effizient ausgenutzt wird.

In der Bildunterschrift wird der Einsatz gestalterischer Mittel und bei system-initiierten Veränderungen auch das durch sie verfolgte Ziel kommentiert. Dazu werden Sichtbarkeitsinformationen und Informationen über die Darstellungsweise von Objekten ausgenutzt.

Anhand der Sichtrichtung kann bestimmt werden, welche Objekte in der bildlichen Darstellung sichtbar sind, bzw. durch welche anderen Objekte sie — auch teilweise — verdeckt werden. Da diese Berechnung in einem interaktivem System zu zeitaufwendig ist, werden für eine Reihe von Blickwinkeln diese *Sichtbarkeitsinformationen* im voraus berechnet. Dazu werden im Abstand von 45° Strahlen in das Modell geschickt und analysiert, in welcher Reihenfolge die Strahlen auf Objekte treffen. Die relative Sichtbarkeit eines Objektes ergibt sich als Quotient aus der Zahl der Strahlen, die das Objekt an erster Stelle treffen und der Zahl der Strahlen, die das Objekt insgesamt treffen. Neben der relativen Sichtbarkeit werden die relative Größe und die Position von allen Objekten bestimmt. Die Sichtbarkeitsinformationen werden zum einen ausgenutzt, um für verdeckte Objekte zu bestimmen, durch welche anderen Objekte sie überlagert werden, um diese eventuell semi-transparent darzustellen. Zum anderen kann auf sie auch zur innerhalb von Templates zur Kommentierung der Sichtbarkeit zugegriffen werden. Die Templates in Abbildung 5 nutzen diese Sichtbarkeitsinformationen.

Neben der interaktiven Änderung der Sichtrichtung können die *Präsentationsvariablen* von Objekten durch den Benutzer selbst oder das System angepaßt werden. Präsentationsvariablen betreffen die Darstellungsweise (z.B. Farben, Texturen, Transparenz und Glanz) von Objekten. Durch Veränderung der Präsentationsvariablen können einzelne Objekte farblich hervorgehoben oder semi-transparent dargestellt werden. Derartige Veränderungen können auch auf Objektklassen angewendet werden. Diese Veränderungen werden in der Bildunterschrift beschrieben.

Besonders wichtig ist in der Bildunterschrift der Verweis auf Veränderungen der Visualisierung, die nicht direkt auf eine Benutzerinteraktion zurückgehen, sondern implizit ausgelöst wurden. So wird, nachdem der Benutzer eine textuelle Erklärung zu einem dargestellten Objekt angefordert hat, das zugehörige Objekt in der graphischen Darstellung farblich hervorgehoben, eventuell leicht vergrößert und es werden verdeckende Objekte durchscheinend dargestellt. Obwohl diese Anpassung sehr nützlich ist, sind die auftretenden Veränderungen eventuell schwer interpretierbar und müssen deshalb in der Bildunterschrift erläutert werden. So kommentieren die letzten beiden Sätze der Bildunterschrift in Abbildung 1 Veränderungen der Darstellung, die vom System initiiert wurden, um den Muskel *Pars lacrimalis* im Bild hervorzuheben. Diese Veränderungen der Präsentation des Muskels *Pars lacrimalis* wurden durch das System, infolge der Anforderung einer erweiterten Annotation zu diesem Muskels durch den Betrachter, ausgelöst.

Die Generierung der Bildunterschrift mit Hilfe des template-basierten Ansatzes erfordert keine großen Laufzeiten, da keine vollständige Satzgenerierung erfolgen muß. Dadurch ist der Einsatz der Bildunterschriften in einem interaktiven System möglich. Andererseits zeigen gerade die letzten beiden Sätze der Bildunterschrift in Abbildung 1 die Grenzen des template-basierten Ansatzes, der oftmals sehr monotone Texte erzeugt. So könnten beide Sätze zusammengefaßt werden, da sie Eigenschaften des gleichen Objektes beschreiben. Eine bessere Variante könnte lauten: *The Pars lacrimalis, which is colored strong red, is solid and slightly enlarged.*

#### 4 Bildunterschriften in interaktiven Systemen

Der Einsatz von Bildunterschriften in einem interaktiven System führt zu ganz neuen Möglichkeiten, wie der Anpassung der Bildunterschrift an den Benutzer, aber auch zu neuen Problemen wie der Wahrung der Konsistenz zwischen Bild- und Bildunterschrift. Wir werden beide Punkte nacheinander diskutieren.

**Konfiguration von Bildunterschriften:** In der Bildunterschrift können eine Reihe von Veränderungen reflektiert werden: Die Sichtrichtung auf das 3D-Modell, selektive Größenveränderungen, Veränderungen der Transparenz von einzelnen Objekten und Objektklassen und farbliche Hervorhebungen. Die Beschreibung all dieser Veränderungen kann lang und unübersichtlich werden. Daher ist eine Priorisierung der Informationen nützlich. Zu diesem Zweck gibt es eine Konfigurationsmöglichkeit, in der ein besonderes Interesse an einzelnen Objekten oder an der Erläuterung des Einsatzes bestimmter gestalterischer Mittel angemeldet werden kann. Auf diese Weise kann die Inhaltsauswahl und vor allem die Ausführlichkeit der generierten Bildunterschriften interaktiv beeinflußt werden.

**Konsistenz zwischen Bild- und Bildunterschrift:** Die Navigation des Benutzers innerhalb von 3D-Darstellungen kann dazu führen, daß die Bildunterschrift nicht mehr der aktuellen Bildschirmdarstellung entspricht — die Konsistenz zwischen Bild- und Bild-

unterschrift wird verletzt. Wir bezeichnen solche Bildunterschriften als *ungültig*. Dieser Umstand erfordert es, die Bildunterschriften zu aktualisieren oder zumindestens die ungültigen Bestandteile zu kennzeichnen. Daher wird die Komponente zur Generierung der Bildunterschriften über alle Veränderungen an der Bild- oder Textdarstellung informiert. Diese analysiert die Veränderung darauf hin, ob die Bildunterschrift unvollständig oder gar ungültig geworden ist und kennzeichnet die ungültigen Bestandteile farblich. Zwei Strategien können angewendet werden, um festzulegen, in welchen Situationen die (ungültige) Bildunterschrift aktualisiert wird:

Zum einen kann das System die Bildunterschrift bei jeder Veränderung automatisch aktualisieren. Dieser Variante ist problematisch, da der Teil der Illustration, in dem die Bildunterschrift präsentiert wird, einer ständigen Aktualisierung unterliegt, was stark irritieren kann. Zum anderen kann sich das System darauf beschränken, Inkonsistenzen zu kennzeichnen. Für die Aktualisierung der Bildunterschrift wird ein entsprechendes Kommando angeboten, so daß die Aktualisierung interaktiv kontrolliert werden kann. Zwischen diesen Extremen gibt es hybride Varianten, bei dem das System bei gravierenden Veränderungen (z.B. solchen, die sich auf die Ansicht als Ganzes beziehen) automatisch die Bildunterschrift aktualisiert, während aufgrund einer Benutzeranforderung jederzeit eine Aktualisierung veranlaßt wird. Es wurden Algorithmen entwickelt, um solche Inkonsistenzen festzustellen und angemessen darauf zu reagieren. Der Benutzer kann zwischen diesen verschiedenen Strategien wählen.

## 5 Diskussion und Ausblick

Die Generierung von multimedialen Präsentationen, insbesondere von Bild-Text-Kombinationen, ist in einer Reihe von Projekten (z.B. WIP [Wahlster et al., 1993] und COMET [Feiner und McKeown, 1993]) behandelt worden. Viele dieser Projekte (insbesondere die beiden genannten) sind in der Domäne der technischen Dokumentation angesiedelt. Dabei werden Handlungsanleitungen in Form von instruktiven Texten generiert, die in dieser Arbeit nicht betrachtet wurden. Weiterhin werden nur in wenigen Projekten 3D-Graphik und Textgenerierung kombiniert; keines der angeführten Projekte erlaubt ein interaktives Erkunden des dargestellten 3D-Modells.

Deskriptive Bildunterschriften werden in SAGE [Mittal et al., 1995] zur Erklärung von abstrakt graphischen Bildern<sup>1</sup> eingesetzt, wobei die Bedeutung von graphischen Elementen sowohl in Legenden als auch in der Bildunterschrift erklärt wird. Es wird dabei betont, daß durch die verbale Erläuterung komplexere Gestaltungsmittel benutzt werden können, die ohne Bildunterschrift unverständlich bleiben würden. Trotz dieser ähnlichen Motivation unterscheidet sich unsere Anwendung grundlegend vom SAGE-Projekt, da in der vorliegenden Arbeit Bildunterschriften zu Illustrationen von 3D-Modellen generiert werden.

Der template-basierte Ansatz hat, wie in Abschnitt 3 gezeigt, eine Reihe von Schwächen, die zu monotonen und langen Texten führen können. Werden Templates in verschiedenen Sprachen bereitgestellt, steigt die Anzahl der notwendigen Templates schnell, so daß sie schwer zu warten sind [Reiter und Mellish, 1993]. Aus diesen Gründen wurde eine Schnittstelle zu einem multilingualen Generierungssystem definiert. Dabei wird zuerst

---

<sup>1</sup>[Strothotte und Strothotte, 1997] verwenden den Begriff *abstrakt graphische Darstellungen* für solche Bilder, die für das menschliche Auge unsichtbare Attribute einer Szene darstellen, um Diagramme, Tabellen, etc. von Darstellungen rein konkreter Objekte abzugrenzen.

die interne Repräsentation der Objekte und Objektklassen, sowie die Sichtbarkeitsinformationen des ZOOM ILLUSTRATORS, in eine Wissensrepräsentation überführt. Wie im vorliegenden Prototyp auch, wird der Textgenerator über alle benutzer- und systeminitiierten Veränderungen informiert. Anhand dieser Informationen können dann multilinguale Bildunterschriften generiert werden. Weiterhin können durch diesen Ansatz die Annotationen, die bisher nur als fester Text in mehreren Versionen vorliegen, ebenfalls generiert werden – vorausgesetzt der Inhalt dieser Annotationen wird ebenfalls in der Wissensbasis repräsentiert.

Eine weitere interessante Anwendung der in der vorgestellten Arbeit erreichten Ergebnisse ist es, die erzeugten Bildunterschriften zur inhaltlichen Beschreibung bestimmter, durch den Benutzer ausgewählter, Stellen der interaktiv erzeugten Präsentationen zu nutzen (Bookmarks) oder zu Bildschirmabzügen von Einzelbildern automatisch eine aussagekräftige Bildunterschrift zu erzeugen.

Der Übergang von Bildunterschriften zur sprachlichen Begleitung von Bildern ist aus verschiedenen Gründen wünschenswert. Hervorzuheben ist die Möglichkeit, Bild und Sprache parallel wahrzunehmen. Für diesen Übergang sind noch eine Reihe weiterer Untersuchungen erforderlich. Diese betreffen insbesondere folgende Punkte:

- **Zeitliche Koordination:** Bei Präsentationssystemen ist es verlockend, verstärkt graphische Animationen einzusetzen. Dabei ist die zeitliche Koordination zwischen den Szenen der Animation bzw. den einzelnen Aktionen innerhalb einer Szene und dem Kommentar erforderlich. Zu beachten ist, daß die Szenenlänge bzw. die Aktionsdauer die Länge des sprachlichen Kommentars diktiert. Wird ein längerer sprachlicher Kommentar als nötig erachtet, müssen geeignete Verzögerungstaktiken in der graphischen Präsentation eingesetzt werden.
- **Kohärenz zwischen Kommentaren:** Sprachliche Kommentare stehen immer im Zusammenhang zueinander. Dies gilt für Kommentare zu Bildfolgen in wesentlich stärkerem Maße als für Bildunterschriften zu Einzelbildern. So braucht z.B. die Sichtrichtung nur von Zeit zu Zeit beschrieben werden.
- **Dramaturgien:** Um monotone Präsentation zu vermeiden, sollten formale Beschreibungen von Dramaturgien verwendet werden, auf welche bei der Generierung der multimedialen Präsentation zugegriffen wird. Diese sollten Einfluß auf die Inhaltsauswahl, die Zuordnung der Inhalte auf die einzelnen Präsentationskanäle und ihre Koordination nehmen. Durch solche Dramaturgien können beispielsweise die Sprechgeschwindigkeit und ähnliche Parameter an die inhaltlichen Höhepunkte der Präsentation angepaßt werden. Die graphisch Darstellung muß ebenfalls angepaßt werden, wobei bisher unklar ist, mit welchen Mitteln dies geschehen kann. Ferner ist zu untersuchen, wie sprachliche Kommentare und Graphik in Einklang mit Begleitmusik gebracht werden können. Diese wird für wichtig erachtet, um die Aufmerksamkeit des Betrachters aufrecht zu erhalten und die Dramaturgie zu untermauern.

Als Ausgangspunkt für die Lösung dieser Fragestellungen kann eine linguistische Analyse der Kommentare in Lehrvideos dienen, die dort in sehr effektiver Weise die dargestellten Inhalte ergänzen.



## 6 Zusammenfassung

Bildunterschriften bilden einen ersten Schritt zur sprachlichen Kommentierung von Bildern in multimedialen Präsentationen. Der Einsatz von Bildunterschriften macht es möglich, komplexere Bilder zu generieren, die ohne eine sprachliche Beschreibung unverständlich bleiben würden. In der vorgestellten Arbeit werden erstmalig Bildunterschriften in *interaktiven* Systemen eingesetzt. Die generierten Bildunterschriften erläutern komplexe räumliche Zusammenhänge und beschreiben dazu die Sicht auf ein 3D-Modell. Dabei wird die Konsistenz zwischen der bildlichen Darstellung und der textuellen Erklärung gewahrt, indem ungültige Passagen gekennzeichnet werden und auf Anforderung aktualisiert werden. Weiterhin wird die Möglichkeit der Anpassung der zu generierenden Bildunterschrift an Benutzerpräferenzen eingeräumt.

## Literatur

- [Bernard, 1990] Bernard, R. M. „Using Extended Captions to Improve Learning from Instructional Illustrations“. *British Journal of Educational Technology*, 21(3):215–225, 1990.
- [Feiner und McKeown, 1993] Feiner, S. K. und McKeown, K. R. „Automating the Generation of Coordinated Multimedia Explanations“. In: Maybury, M. T., Hrsg., *Intelligent Multimedia Interfaces*, Seite 117–138. AAAI Press, Menlo Park, CA, 1993.
- [Gombrich, 1984] Gombrich, E. H. *Bild und Auge. Neue Studien zur Psychologie der bildlichen Darstellung*. Klett-Cotta, Stuttgart, 1984.
- [Mittal et al., 1995] Mittal, V. O., Roth, S., Moore, J. D., Mattis, J., und Carenini, G. „Generating Explanatory Captions for Information Graphics“. In: Mellish, C. S., Hrsg., *Proc. of the 14th International Joint Conference on Artificial Intelligence (IJCAI-95)*, Seite 1276–1283, Montréal, Québec, Canada, August 20–25 1995. Morgan Kaufmann Publishers, San Mateo, CA, USA.
- [Preim et al., 1996] Preim, B., Ritter, A., und Strothotte, T. „Illustrating Anatomic Models – A Semi-Interactive Approach“. In: Höhne, K.-H. und Kikinis, R., Hrsg., *Visualization in Biomedical Computing: 4th International Conference (VCB'96)*, Seite 23–32, Hamburg, Germany, September 22-25 1996. Springer Verlag, Berlin, Heidelberg, New York.
- [Preim et al., 1997] Preim, B., Raab, A., und Strothotte, T. „Coherent Zooming of Illustrations with 3D-Graphics and Text“. In: *Proc. of Graphics Interface '97*, Seite 105–113, Kolowna, BC, May 21–23 1997.
- [Reiter und Mellish, 1993] Reiter, E. und Mellish, E. „Optimizing the Costs and Benefits of Natural Language Generation“. In: *Proc. of the 13th International Joint Conference on Artificial Intelligence (IJCAI'93)*, Seite 1164–1169, Chambery, France, August 28 – September 3 1993.
- [Sommerfeld, 1997] Sommerfeld, T. „Generierung von deskriptiven Bildunterschriften zur Illustration von 3D-Modellen“. Diplomarbeit, Otto-von-Guericke-Universität Magdeburg, Fakultät für Informatik, Institut für Simulation und Graphik, Magdeburg, April 1997.
- [Staubesand, 1988] Staubesand, J., Hrsg. *Sobotta; Atlas der Anatomie des Menschen, Band 2: Brust, Bauch, Becken, untere Extremitäten*. Urban & Schwarzenberg, München, Wien, Baltimore, 1988. 19. Auflage.
- [Strothotte und Strothotte, 1997] Strothotte, C. und Strothotte, T. *Seeing Between the Pixels; Pictures in Interactive Systems*. Springer, Berlin, 1997. im Druck.

- [Strothotte und Wagener, 1997] Strothotte, T. und Wagener, H., Hrsg. *Abstraction in Interactive Computer Visualizations: Exploring Complex Information Spaces*. Springer Verlag, Berlin, Heidelberg, New York, 1997. (erscheint).
- [Wahlster et al., 1993] Wahlster, W., André, E., Finkler, W., Profitlich, H.-J., und Rist, T. „Plan-Based Integration of Natural Language and Graphics Generation“. *Artificial Intelligence*, 63:387–427, 1993.
- [Weidenmann , 1989] Weidenmann, B. „Informative Bilder – was sie können, wie man sie didaktisch richtig nutzen und wie man sie nicht verwenden sollte “. *Pädagogik*, 9:30–33, 1989.