In *Proc. CHI '98,* pages 392–399, Los Angeles, CA, April 18–23, 1998

Visual Task Characterization for Automated Visual Discourse Synthesis

Michelle X. Zhou **Steven K. Feiner**
Department of Computer Science
Columbia University
500 West 120th St., 450 CS Building
New York, NY 10027
+1 212 939 7000
{zhou, feiner}@cs.columbia.edu

ABSTRACT

To develop a comprehensive and systematic approach to the automated design of visual discourse, we introduce a visual task taxonomy that interfaces high-level presentation intents with low-level visual techniques. In our approach, visual tasks describe presentation intents through their visual accomplishments, and suggest desired visual techniques through their visual implications. Therefore, we can characterize visual tasks by their visual accomplishments and implications. Through this characterization, visual tasks can guide the visual discourse synthesis process by specifying what presentation intents can be achieved and how to achieve them.

Keywords

Automated design of graphics, visual discourse, visual task characterization.

INTRODUCTION

Handcrafting effective visual presentations can be extremely expensive and time-consuming, especially if the presentations are to be customized for individual users and situations. Therefore, researchers are developing computational approaches to automating the design process. The work described here addresses the general task of designing *visual discourse*, a series of connected visual displays [32], of which a single display or a series of static displays may be considered special cases.

If an approach is to create effective presentations, it must accomplish a set of presentation *intents* associated with the discourse [3, 21, 23, 8]; for example, it may inform a user about some fact *x* or enable a user to search for information *y*, which we will notate as Inform<?x> or Search<?y>. One design approach, which we are investigating, involves the formulation of a set of rules to select appropriate visual techniques based on the presentation intents. However, it is difficult to formulate general rules that can directly relate high-level presentation intents, such as Inform<?x>, to low-level visual techniques, such as Highlight<?x>.

Based on a variety of visual presentations from different sources (e.g., [22, 26, 27, 10, 28]), and our own work on visual discourse synthesis [32], we believe that there is a level of abstraction midway between presentation intents and low-level visual techniques. We refer to this level of abstraction as a *visual task taxonomy.* Unlike presentation intents, visual tasks *directly* indicate the desired visual effects. For example, the visual task Locate<?x, ?locator> signifies that an object ?x be visually located relative to another object, ?locator. However, visual tasks are *not* visual techniques, as they may imply related visual techniques without specifying exactly *how.* For example, the visual task Focus<?x> may imply that visual techniques such as Enlarge<?x> or Highlight<?x> could be used to focus attention on ?x, but without specifying which technique to use.

On one hand, visual tasks can be grouped together to uniformly describe how to accomplish various presentation intents. On the other, as visual tasks imply desired visual techniques, they enable abstraction beyond the concrete details of lower-level visual techniques, while still maintaining close enough contact with these techniques. Therefore, visual tasks could be considered as *abstracted* visual techniques, which can be achieved by a set of lower-level techniques. Consequently, high-level presentation intents and low-level visual techniques are tied together through visual tasks: a presentation intent could be accomplished by certain visual tasks, which then imply the visual techniques to achieve them.

As visual tasks serve as an interface between presentation intents and low-level visual techniques, they can be characterized along two dimensions: visual accomplishments and visual implications. *Visual accomplishments* specify the presentation intents a visual task could achieve, while *visual implications* indicate the visual techniques a visual task may imply. Using the visual task taxonomy, visual discourse synthesis first tries to interpret the presentation intents using a set of visual tasks; then, based on the implications of these visual tasks, it suggests a set of visual techniques to achieve the presentation intents.

RELATED WORK

To characterize different visual presentations and capture the relationships between presentation intents and visual cues, researchers have developed various taxonomies (e.g., [7, 13, 24, 29, 10]). Unlike our visual task taxonomy, previous taxonomies either broadly categorize the functions of

This research is supported in part by DARPA Contract DAAL01-94-K-0119, the Columbia University Center for Advanced Technology in High Performance Computing and Communications in Healthcare (funded by the New York State Science and Technology Foundation), the Columbia Center for Telecommunications Research under NSF Grant ECD-88-11111, and ONR Contract N00014-97-1-0838.

pictures without examining their specific visual organizations (e.g., [7, 13]), or only describe conceptual relationships between *information-seeking goals* and various visual techniques (e.g., [29, 10]). Of these researchers, only Sutcliffe et al. [24] briefly mention four specific visual tasks (e.g., Highlight and Classify).

For the purpose of automated visual discourse synthesis, researchers have classified various user information-seeking goals (e.g., [3, 21, 8]) or communicative acts (e.g., [1, 17, 23]). Although information-seeking goals or communicative acts serve different design purposes [20], they are similar in the sense that they both specify high-level presentation intents. To bridge the gap between high-level presentation intents and low-level visual techniques, a set of intermediate specifications is usually constructed.

One type of specification, known as *perceptual operators* (e.g., [3, 8]), indicates the perceptual tasks to be performed by the user in a visual environment. There are significant differences between a perceptual operator and a visual task. To achieve presentation intents, perceptual operators emphasize what a *user* needs to do (e.g., *search* for an object), while visual tasks describe what needs to be carried out by a visual *presentation* (e.g., make the object of search be in *focus*). In this sense, the visual task operates one level *below* the perceptual operator to signify the desired visual effects. Whereas perceptual operators are usually employed sequentially, visual tasks can be specified in parallel to achieve multiple visual effects simultaneously.

Closest to our visual task specifications are *visual acts,* such as style strategies in [23], visual operators in [1], and graphic acts in [17]. Like visual tasks, visual acts *directly* specify the visual effects. By our definition, however, visual acts such as Highlight, Zoom, and Include are more like low-level visual techniques than *abstracted* visual tasks. Instead of presenting an abstraction of various visual techniques, these visual acts are usually employed directly to achieve communicative intents, even though some of their parameters (e.g., a Highlight color) may be determined in the application stage. Moreover, since systems that use visual acts deal mainly with real-world objects (e.g., [1, 23]) or existing presentations (e.g., a database of maps in [17]), they focus on visual manipulation instead of visual creation. Therefore, they rarely include or describe visual acts that can be used to create visual presentations from scratch (e.g., visual symbol encoding acts in graph creation). In comparison, our visual task taxonomy presents an abstraction above low-level visual techniques and addresses both visual creation and manipulation.

It is also worth noting that the *pragmatic directives* proposed by Marks [16] are different from any of the taxonomies mentioned above. They more or less directly imply the underlying visual organization or layout (e.g., a particular network path or a hub), but do not imply or suggest any of the low-level attribute encodings. In contrast, our visual task taxonomy covers both high-level visual organization and low-level visual symbol encoding tasks.

VISUAL TASK CHARACTERIZATION

Building on previous work [29, 10, 17, 24], we have derived a taxonomy of visual tasks, shown in Figure 1. Like a visual act, each visual task is described by two parts: an act and a set of arguments to act on. We have refined each type of visual task so that it can be accurately described and effectively employed to achieve presentation intents. These refinements could be interpreted as *particular ways* of achieving the task. For example, the visual task Identify<?x, ?identifier> could be refined using one of the four more specific visual tasks to indicate a particular way of *identifying* an object:

Name: identify ?x using its name
Portray: identify ?x by its portrait (image)
Individualize: identify ?x using its unique attributes
Profile: identify ?x using its profile

We characterize visual tasks along two dimensions: visual accomplishments and visual implications. *Visual accomplishments* describe the type of presentation intents that a visual task might help to achieve, while *visual implications* specify a particular type of visual action that a visual task may carry out.

Visual Accomplishments

Informational visual presentations are usually charged with one of two intents [2, 20]: the presentation is intended either to simply convey a presenter's message to a user or to help the user accomplish certain perceptual tasks such as *search* or *verify*. Based on this distinction, visual tasks could

\# Relational visual tasks
- **Associate**<?x, ?y>
  - Collocate<?x, ?y>
  - Connect<?x, ?y>
  - Unite<?x, ?x-part>
  - Attach<?x, ?x-part>
- **Background**<?x, ?background>
- **Categorize**<?x1, ...,?xn>
  - MarkDistribute<?x1, ..., ?xn>
- **Cluster**<?cluster, ?x1, ..., ?xn>
  - Outline<?cluster>
  - Individualize<?cluster>
- **Compare**<?x, ?y>
  - Differentiate<?x, ?y>
  - Intersect<?x, ?y>
- **Correlate**<?x1, ..., ?xn>
  - Plot<?x1, ..., ?xn>
  - MarkCompose<?x1, ..., ?xn>
- **Distinguish**<?x, ?y>
  - MarkDistribute<?x, ?y>
  - Isolate<?x ?y>
- **Emphasize**<?x-part ?x>
  - Focus<?x-part ?x>
  - Isolate<?x-part ?x>
  - Reinforce<?x-part ?x>
- **Generalize**<?x1, ..., ?xn>
  - Merge<?x1, ..., ?xn>
- **Identify**<?x, ?identifier>
  - Name<?x, ?name>
  - Portray<?x, ?image>
  - Individualize<?x, ?attr>
  - Profile<?x, ?profile>
- **Locate**<?x, ?locator>
  - Position<?x, ?locator>
  - Situate<?x, ?locator>
  - Pinpoint<?x, ?locator>
  - Outline<?x, ?locator>
- **Rank**<?x1, ..., ?xn, ?attr>
  - Time<?x1, ..., ?xn, ?t>
- **Reveal**<?x-part ?x>
  - Expose<?x-part ?x>
  - Itemize<?x-part ?x>
  - Specify<?x-part ?x>
  - Separate<?x-part ?x>
- **Switch**<?x, ?y>

\# Direct visual organizing and
\# encoding tasks
- **Encode**<?x>
  - Label<?x>
  - Symbolize<?x>
    - Quantify<?x>
    - Iconify<?x>
  - Portray<?x>
  - Tabulate<?x>
  - Plot<?x>
  - Structure<?x>
  - Trace<?x>
  - Map<?x>

Figure 1. Visual task taxonomy

be classified into two categories at the highest level: visual tasks that help the presenter to *inform* the user and tasks that *enable* the user to perform visual *exploration* or *computation*. Visual tasks that *inform* could be further distinguished as to whether they help to *summarize* or to *elaborate* information. Similarly, visual tasks that enable a user to *explore* or *compute* could be further classified by whether they aid the user to *search* for a particular object, to *verify* a fact, or to *differentiate* or *sum* different values.

As listed in Figure 2, many visual tasks could help to achieve different presentation intents at once in different ways [13]. For example, the visual task Compare<?x, ?y> could aid in accomplishing the intent Inform<?x> by indicating the similarity between ?x and ?y (Intersect<?x, ?y>), and could also help to achieve the intent Verify<?x> by showing a difference (Differentiate<?x, ?y>). To decide which visual tasks may be used to achieve what presentation intent, we pair presentation intents and corresponding information categories together [9] (e.g., Verify<object-composition>). Then we infer how the specific presentation intent could be accomplished *visually*, based on both rhetorical and visual design principles [22, 18, 15, 26, 27, 19, 28]. For example, Verify<object-composition> requires displaying an object and its components. Thus, the visual task Reveal<?x> could facilitate perceptual verification.

In addition to information categories, many other factors, such as the audience or presentation environment, will also affect the mapping between presentation intents and visual tasks. For example, informing a nurse about a patient's postoperative status may require accomplishing visual tasks such as Locate<devices, patient-body> (locate medical devices in context of the patient's body) and Itemize<drugs> (list all administered drugs). In contrast, informing a doctor about the same patient may involve visual tasks such as Trace<operation-events> (order events during the operation temporally) and Categorize<drug1, ... purpose> (categorize all administered drugs by purpose).

Therefore, visual task classification by visual accomplishments based on information categories alone is incomplete. Other factors, such as the user model, must be taken into account to establish a comprehensive mapping between presentation intents and visual tasks. Here, we focus on discussing the general *domain-independent* connections between presentation intents and visual tasks, and hope that such a broad characterization will serve as a starting point in visual discourse synthesis. By accounting for the factors mentioned above, more accurate mappings and refined characterizations for specific domains can then be formulated based on these general guidelines (see Section **EXAMPLES**).

Thus far, we have roughly characterized various visual tasks by their accomplishments. Next we introduce a finer-grained categorization of visual tasks by their visual implications.

Visual Implications

Cognitive psychologists have conducted extensive studies to understand human visual perceptual behavior and reveal how visual cues can affect or direct perception (e.g., [25, 6, 12]). Based on these studies, a set of principles for visual perception and cognition has been formulated. From the standpoint of visual discourse synthesis, we have summarized three types of visual perception and cognition principles: visual organization, visual signaling, and visual transformation. The *visual organization* principle suggests how people visually organize the world and perceive it as a coherent whole (e.g., [25, 6, 12]). The *visual signaling* principle explains how people tend to interpret visual cues and infer their meanings (e.g., [6]). The *visual transformation* principle explains how people switch attention and adapt to visual changes (e.g., [31]).

Directed by these principles, we could categorize various visual tasks by their visual implications: whether they imply certain types of *visual organization*, certain ways of *visual signaling*, or certain paths of *visual transformation*. Based on these implications, a visual discourse synthesis system may use visual tasks to sketch the visual organization of the presentation, and to search for appropriate visual signaling or transformation techniques with which to encode information in a new presentation or modify an existing one. Moreover, we could formulate partial plans in advance for each type of visual task, using lower-level visual techniques [32]. For example, if we know that the visual task Focus implies attracting attention, and that visual techniques such as Highlight or Enlarge attract visual attention, then these techniques could become different subplans to achieve Focus.

Visual Organization

Visual organization could be described from several aspects such as overall pattern or sequence of the organization (e.g., [30]), preattentive features of the organization (e.g, [25]), and various visual relationships among visual elements (e.g., [11, 12]). Generalizing from this previous work, we have compiled a comprehensive set of features to describe various aspects of visual organization: visual grouping, visual attention, visual sequence, and visual composition.

Visual grouping is concerned with how visual patterns

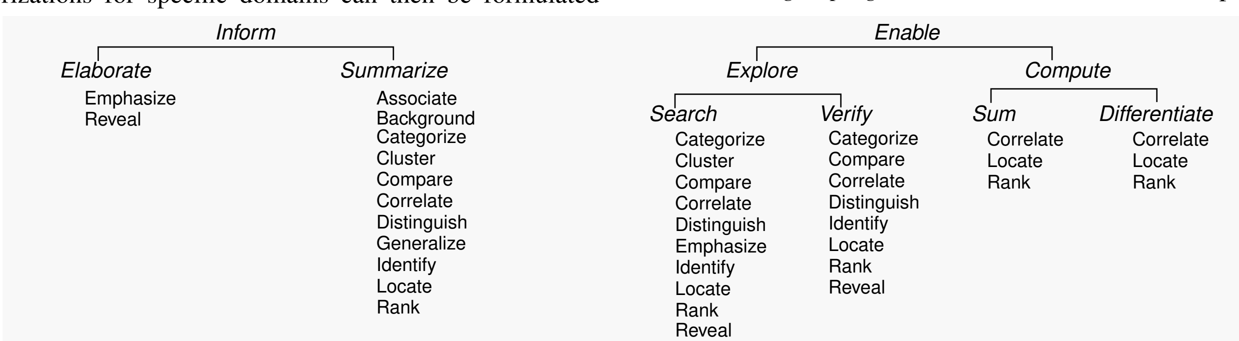

Figure 2. High-level presentation intents (*italic*) and visual tasks (roman) that achieve them

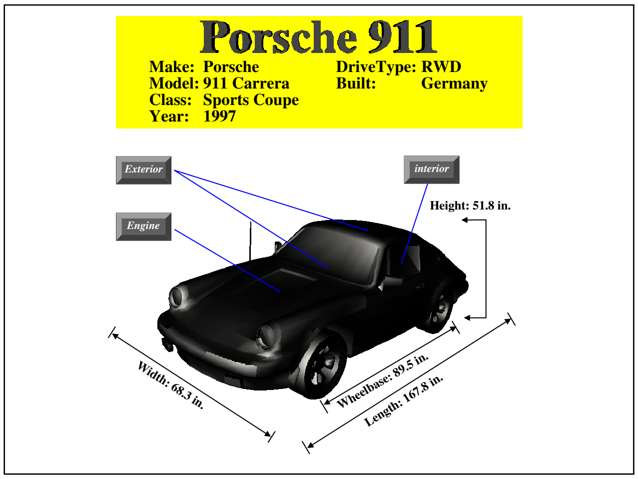

Figure 3. Present information about a car to a customer

could be grouped together so people can perceive them as groups. *Visual attention* asserts what distinct features enable people to perceive them preattentively and effortlessly. *Visual sequence* indicates how visual patterns influence people to scan presentations in a particular order, while *visual composition* focuses on explaining different ways of composing lower-level visual elements into more complex visual structures. Next, we describe how various visual tasks are characterized by these four aspects.

Visual Grouping. According to gestalt psychology, people group different visual elements together by certain patterns: proximity, similarity, continuity, and closure [19].

Proximity states that objects that are spatially close will be perceived as being together. For example, Figure 3 presents information about a car to a customer. Below the presentation's title, it is easy to perceive two groups corresponding to two columns: make, model, class and year are in one group, and the rest in another. The visual task Cluster can be used to indicate this grouping. Based on their semantics, these visual tasks imply visual grouping by proximity:

Associate: associated objects perceived as a group by proximity
Cluster: spatial clusters perceived as groups by proximity
Locate: locator and objects being located are grouped by proximity

Similarity asserts that objects that are presented with similar visual appearance will be perceived as a group. In Figure 4, sales centers are encoded as dark spheres and branches as light cones. At first glance, two groups emerge: the spheres and the cones. This particular grouping is accomplished through the visual task MarkDistribute<?centers, ?branches>, which *categorizes* various objects by visual appearance. Several visual tasks enforce grouping by similarity to achieve certain visual effects:

Categorize: objects in the same category grouped by similarity
Cluster: objects in the same cluster grouped by similarity
Distinguish
  MarkDistribute: similarity helps to distinguish groups

Continuity suggests that people tend to group objects and their continuations together as a whole. Figure 5 displays engine data in the form of a pull-down menu below the *Engine* button. A pull-down menu can be interpreted as a *continuation* of a button that reveals more information when the button is pressed. In this case, the engine button and its pull-down menu continuation are easily perceived as a group. The visual task Itemize<engine-data>, a special case of Reveal<engine-data>, precisely describes the intended grouping. These visual tasks imply visual grouping by continuity:

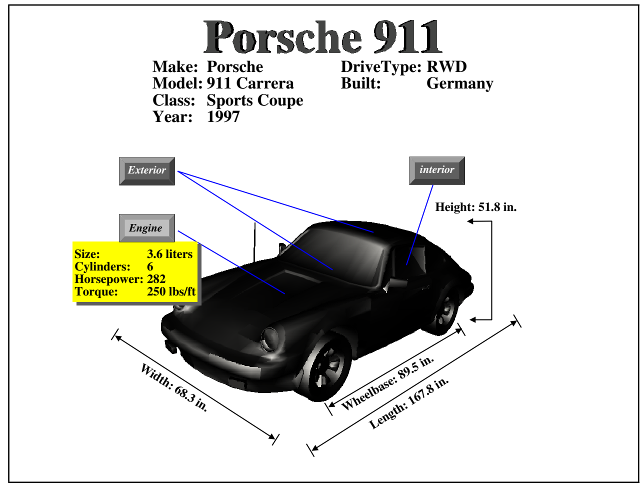

Figure 5. Reveal information about a car's engine

Associate
  Connect: two objects are each other's continuation
  Attach: attachment is a continuation of another object
Locate
  Outline: a spatial or temporal extent marked as a continuation
Reveal
  Expose: internals depicted as spatial continuation
  Itemize: set members depicted as conceptual continuation
  Specialize: specifications depicted as conceptual continuation

Closure implies that people tend to group things together as a coherent whole when they are in a closed form. As depicted in Figure 4, the map used as the locator (Locate<nodes, map>) presents a closed form to group everything together as a network so they are not perceived as unre-

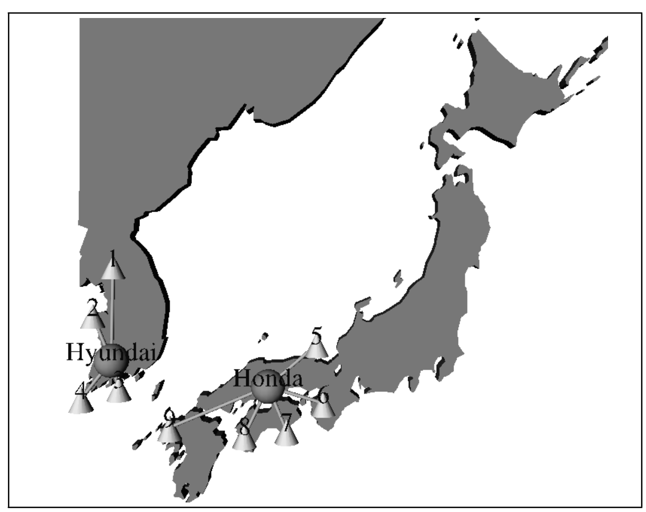

Figure 4. Car sales network

lated entities. Other visual tasks suggesting closure are:

Cluster
  Outline: visual clusters indicated by closed form
Locate: locator presents objects in a closed form

Besides proximity, similarity, continuity, and closure, *symmetry* groups objects by a sense of "good form" [11]. However, since the definition of "good form" may depend on factors such as the viewer's visual literacy, we do not discuss it here.

Visual Attention. Cognitive psychology studies have shown that people are usually drawn to special visual features when they gaze at visual presentations (e.g., [25, 6]). As the process of recognizing these features does not require conscious attention [25], it is important to know what types of visual tasks imply such *preattentive* features. Using this knowledge, presentation systems could employ these visual tasks to achieve related presentation intents, and use appropriate visual techniques to accomplish these tasks.

Since we focus only on visual features independent of their semantics, our visual tasks are concerned only with *syntactic emphasis* as defined by Goldsmith [6]. Goldsmith summarizes eight important visual factors that help syntactic emphasis: color, position, size, isolation, complexity, tonal contrast, directionality, and implied motion.

We examine whether a visual task implies exploiting any of these eight factors to attract or direct attention. One visual task of Figure 3 is to *reinforce* the general information displayed at the top (Reinforce <general-info>). Directed by the color and tonal contrast, the user readily focuses his attention on the highlighted information. In other words, the visual task Reinforce implies attracting and directing attention. The visual tasks related to visual attention are:

Cluster
  Outline: spatial clusters marked by outline attract attention
Distinguish
  Isolate: spatially or perceptually isolated objects attract attention
Emphasize: emphasized objects attract attention
Locate
  Pinpoint: pinpointed object attracts attention
  Outline: marked spatial or temporal extent attracts attention

Visual Sequence. Although people can perceive multiple visual features simultaneously, cognitive studies assert that some pattern recognition and interpretation still requires *successive* processing [25]. *Visual sequence* therefore becomes another important factor in visual organization. In other words, if visual elements are organized in the right sequence, the resulting presentation will guide a viewer to process information efficiently. In contrast, poorly organized presentations will greatly impair a viewer's performance [30].

There are two types of visual sequence, depending on what visual features produce the order: spatial sequence and perceptual sequence. *Spatial sequence* indicates how people are intended to successively scan a presentation based on its elements' *position*s, whereas *perceptual sequence* indicates how people order patterns or features by their visual *appearance*. Winn and Holliday [30] have summarized desired visual sequences used in charts, graphs and diagrams. Goldsmith's experiments [6] have also indicated how certain visual patterns guide ordered perception.

Based on their results, we recognize a set of visual tasks that imply certain visual sequences in presentations. For example, one of the visual tasks to be carried out in Figure 3 is to use the general car description (e.g., make and model) to *identify* the car. In this case, the description is encoded as the presentation's title and placed at the top. This arrangement already implies a visual sequence: read the description first, then the rest. In fact, all types of Identify will imply a similar sequence: read the *identifier* first, then the *object* being identified. Likewise, all types of Emphasize tasks also suggest a particular visual sequence: read the emphasized objects first, then the rest. These visual tasks imply a visual sequence:

Emphasize: emphasized objects and the rest are ordered
Identify: identifier and objects being identified are ordered
Rank: objects are ordered by rank

Visual Composition. The last factor that affects visual organization is visual composition. *Visual composition* indicates how different visual elements are pieced together to form a coherent whole based on their relationships. Visual tasks express these relationships and eventually determine their corresponding spatial and perceptual relationships. For example, the visual task Expose is carried out in Figure 6 to reveal a car's engine. The *constituent* relationship between the car and its engine is conveyed by showing the engine spatially inside the car. Similarly, other types of Reveal also imply different types of spatial composition (e.g., Itemize<engine-data> in Figure 5). Not only does Reveal imply certain ways of composing visual elements, but visual tasks such as Identify, Associate, and Correlate also suggest various types of visual composition:

Associate: associated objects imply being placed together visually
Correlate: correlated objects imply composite visual structure
Identify: object and its identifier imply a particular arrangement
Reveal: objects being revealed imply visual composition

Visual Signaling

As abstracted visual techniques, visual tasks usually *imply* desired visual structures or encodings without explicitly naming them. Nonetheless, some visual tasks may *directly* specify the desired visual structures or encodings at

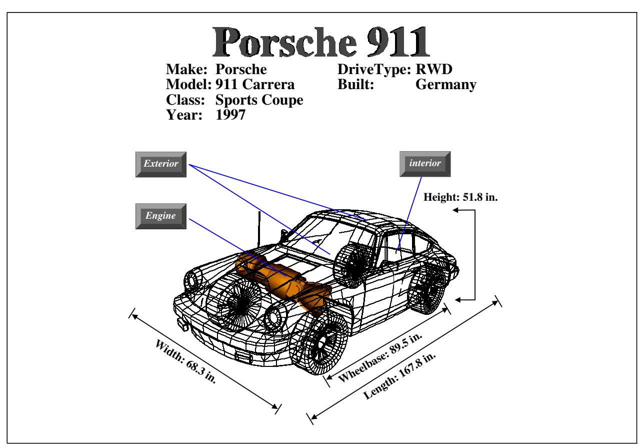

Figure 6. Reveal a car's engine

the user or application's request (e.g., *chart* these data). Whereas visual organization describes the *framework* of the presentation, *visual signaling* indicates the more concrete visual *contents* of the presentation. For example, visual tasks may specify that information be presented *symbolically* (e.g., the car model in Figure 3) rather than *textually* (e.g., the engine table in Figure 3), or in a *cartogram* (e.g., Figure 4) instead of a *structure diagram* (e.g., Figure 3).

Based on the level of detail, we divide visual signaling into two types: visual structuring and visual encoding. *Visual structuring* specifies the type of display [7, 14], while *visual encoding* emphasizes visual symbol encoding, such as using visual variables (e.g., color, size, and shape) to encode an object or its attributes.

Visual Structuring. We have combined several visual display taxonomies [7, 30, 11, 14] to categorize the displays in which we are interested. Based on this display categorization, we recognize five types of visual tasks: Tabulate, Plot, Structure, Trace, and Map. They imply five types of visual display:

Tabulate: implies organizing information in a tabular chart

Plot: implies organizing quantitative information in conventional graphs, such as a bar graph or line graph.

Structure: implies organizing information in a structure diagram by showing the structures of the objects

Trace: implies organizing information in a process diagram or time chart by temporal order.

Map: implies organizing information by geographic location

Visual Encoding. In a visual representation, an object can either be named in a textual format, encoded by a visual variable, or depicted by a realistic image. We refer to these three types of encoding as *labeling* an object, *symbolizing* an object, and *portraying* an object, respectively. Thus, three visual tasks Label, Symbolize, and Portray correspond to three different visual encodings. In addition, Symbolize could be further refined to Quantify or Iconify an object, depending on what visual variable is involved. For example, Quantify implies that the object may be encoded effectively using the visual variable *size*, but not by *shape* or *color* [2, 4], while Iconify implies that the object may be encoded by shape or color [14].

Visual Transformation

To ensure visual discourse coherence, visual transformation cannot be ignored [32]. Visual transformation can help integrate new information into the current display to achieve *unity*; or gradually transform the current display into a new one to achieve *continuity*. We need to identify which visual tasks imply visual transformation, and subsequently, determine how to achieve these transformations.

We recognize two types of visual transformation: visual modification and visual transition. *Visual modification* asserts that the current display has been modified either to achieve a new task or accommodate a new piece of information. *Visual transition*, on the other hand, indicates visual effects [31] that bridge a gap between two completely different presentations (e.g., a cross-fade).

Examining our visual tasks, we are able to determine that some imply visual modification, and others visual transition. For example, various types of Reveal tasks indicate visual modifications by adding more detailed information to the current display (e.g., Itemize), while tasks such as Generalize may suggest visual modifications by merging different components together. As shown in the following table, only the visual task Switch implies a visual transition.

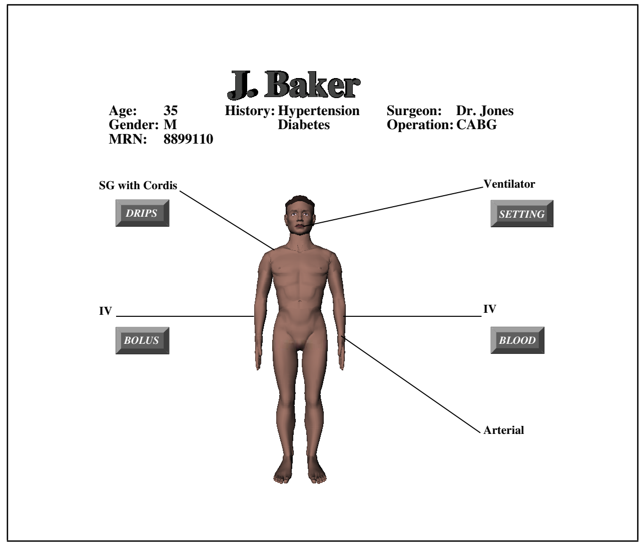

Figure 7. Present a patient's information to a nurse

Emphasize: may imply transformation by modifying emphasized objects

Generalize: implies transformation by merging components

Reveal: implies transformation by adding new information

Switch: implies transition by switching from one scene to another

EXAMPLES

Figure 7 *summarizes* for nurses a patient's status after a coronary artery bypass graft (CABG) operation, while Figure 9 *summarizes* and *elaborates* for doctors a patient's post-operative status. Using these two examples, we demonstrate how IMPROVISE formulates visual tasks based on their visual accomplishments, and how it determines visual techniques based on their visual implications. We focus here on explaining the use of visual tasks in visual discourse synthesis, having described IMPROVISE's inference mechanism in our previous work [32].

Present Patient Information to a Nurse

As shown in Figure 8, the first task formulated by IMPROVISE is to Structure all patient information together, since our informal design studies [5] indicated that nurses prefer all relevant information be placed near a patient's physical body and displayed at once (rather than successively revealed). This task could be directly specified as a user directive, or inferred from a set of rules, as it is in IMPROVISE:

If: Summarize<?info> & Audience<Nurse>
then: Arrange<?info, ?structure> & IsPhysical<?structure>
If: Arrange<?info, ?structure> & IsPhysical<?structure>
then: Structure<?info>

Based on domain information, two more tasks are formulated: using demographics information to *identify* other information, and using the patient's body to *locate* all medical devices. In particular, while the demographics provides a

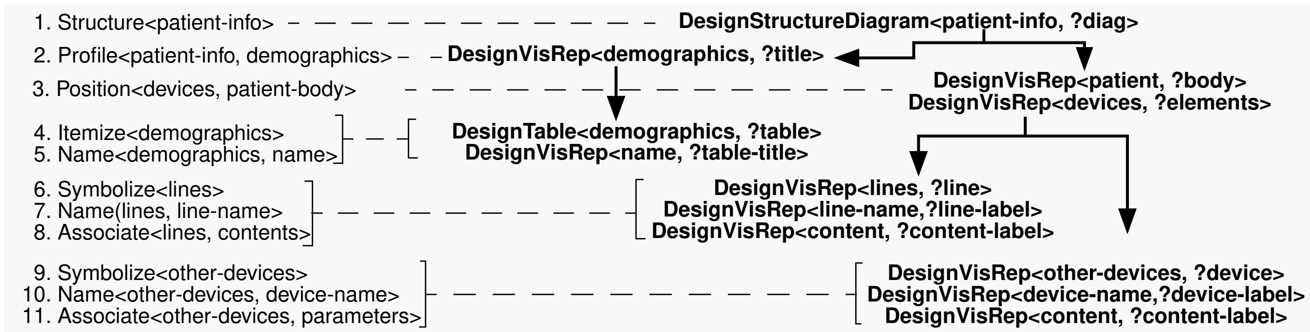

Figure 8. Visual tasks and their corresponding visual techniques used in designing Figure 7

profile of the patient information (task 2 in Figure 8), the medical devices are *positioned* relative to the body (task 3). As IMPROVISE continues, additional visual tasks are formulated to more concretely describe the intended visual relationships among different components (tasks 4–11). Based on user preference, tasks 4–5 involve *itemizing* all demographics information and using the patient name to *name* it. As both infusion lines and other devices must be physically located, tasks 6–11 specify the desired *symbolic* representations for all medical devices that should also be *named* and *associated* with their respective contents or parameters.

Once the visual tasks are formulated, IMPROVISE tries to determine visual techniques for accomplishing them. As mentioned in the last section, Structure implies constructing a structure diagram (DesignStructureDiagram in Figure 8), which is capable of conveying a physical structure (e.g., the patient body) and expressing the spatial coordinates for related objects (e.g., medical devices) [30, 14]. DesignStructureDiagram produces an "empty" structure diagram, and other visual tasks help refine it. Task 2 (Profile) indicates that the demographics become the *title* of the diagram as an identification, while task 3 (Position) makes the patient body become the *core* of the diagram, about which other information can be located. Based on visual tasks 6–11, the design of the structure diagram becomes more complete: Every medical device (e.g., SG with cordis) is symbolized using a line, identified by its textual name, and associated with its contents (e.g., DRIPS button) or parameters.

Present Patient Information to a Doctor

Unlike nurses, doctors in this situation prefer to see events arranged along a time line (Trace<events>). But similar to the nurse's example, the demographics information is used to Identify other information. In this particular example, the presentation should also depict the correlation among the event's time, location, and tasks performed (Correlate<?time, ?location, ?tasks>) at the doctor's request. Moreover, to *elaborate* the patient's bypass status, IMPROVISE formulates three tasks: Outline<status-time>, Itemize<status>, and Focus<bypass-time>. Not only must the presentation communicate the abnormal bypass time qualitatively (e.g., *prolonged*), but it must also Specify the time value quantitatively (e.g., *3 hrs.*). Based on these tasks and object relationships, IMPROVISE is also able to infer that the bypass event itself should also be in focus (Focus<bypass>) as the status is *part of* the bypass event.

While the task Trace<events> implies using a time chart (DesignTimeChart) to organize all events, task Correlate refines the chart by providing its axes (Figure 9). Furthermore, all events are encoded as boxes and placed on the chart. To Itemize the bypass status and Focus on the bypass time, all status information is listed in a panel with bypass time information highlighted. The bypass event block is highlighted and the status information is outlined along the time axis to accomplish Focus<bypass> and Outline<status-time>. Furthermore, the bypass time is also stressed quantitatively by using the time axis to achieve Specify<bypass-time>.

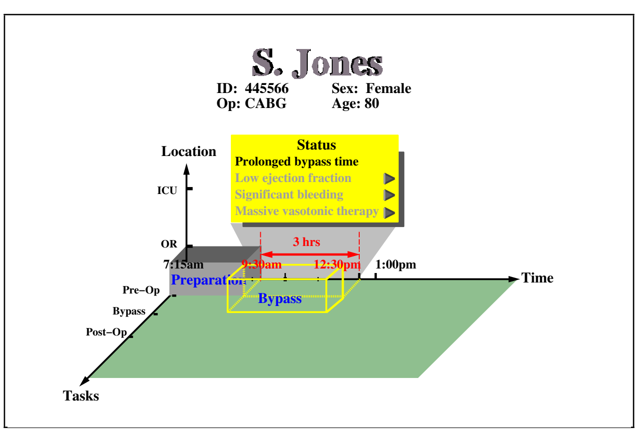

Figure 9. Present a patient's information to a doctor

CONCLUSIONS AND FUTURE WORK

We have identified and presented a set of visual tasks that can serve as an interface between high-level presentation intents and low-level visual techniques. This set of tasks is summarized based on our careful analysis of a wide range of visual presentations (e.g, [26, 27, 10, 28]) and integration of both rhetorical and visual design theories (e.g., [22, 18, 30, 15, 11, 19]). Moreover, these task specifications are being used in our experimental visual discourse synthesis system IMPROVISE to test their practical value.

To use visual task specifications in automated visual discourse synthesis, we have characterized them by their visual accomplishments and implications. In the course of synthesis, a presentation system can formulate a set of visual tasks based on their accomplishments to achieve presentation intents, then select the appropriate visual techniques based on their implications to achieve these tasks. Two different examples generated by IMPROVISE illustrate the visual-task-oriented synthesis process.

To construct a complete visual discourse synthesis system, visual task characterization must be combined with other areas of research, such as data characterization and visual principle modeling [20]. Formal user studies will also be essential to provide a more comprehensive and accurate evaluation of the current set of visual tasks. For example, we would like to determine how well users can employ the visual task descriptions to describe visual presentations, and the range of presentations they can describe.

REFERENCES

[1] E. Andre and T. Rist. The design of illustrated documents as a planning task. In M. Maybury, editor, *Intelligent Multimedia Interfaces*, chapter 4, pages 94–116. AAAI Press/The MIT Press, Menlo Park, CA, 1993.

[2] J. Bertin. *Semiology of Graphics*. Univ. of Wisconsin Press, Madison, WI, 1983. (trans. by W.J. Berg).

[3] S. Casner. A task-analytic approach to the automated design of graphic presentations. *ACM Trans. on Graphics*, 10(2):111–151, 1991.

[4] W. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79:531–554, 1984.

[5] M. Dalal, S. Feiner, K. McKeown, D. Jordan, B. Allen, and Y. alSafadi. MAGIC: An experimental system for generating multimedia briefings about post-bypass patient status. In *Proc. 1996 AMIA Annual Fall Symp*, pages 684–688, Washington, DC, October 26-30 1996.

[6] E. Goldsmith. *Research Into Illustration: An Approach and A Review*. Cambridge University Press, Cambridge, 1984.

[7] B. Hunter, A. Crismore, and P. Pearson. Visual displays in basal readers and social studies textbooks. In D. Willows and H. Houghton, editors, *The Psychology of Illustration: Instructional Issues*, volume 2, chapter 5, pages 116–158. Springer-Verlag, New York, 1987.

[8] E. Ignatius, H. Senay, and J. Favre. An intelligent system for task-specific visualization assistance. *J. of Visual Lang. and Computing*, 5:321–338, 1994.

[9] P. Johnson, H. Johnson, R. Waddington, and A. Shouls. Task-related knowledge structures: Analysis, modelling, and application. In D. Jones and R. Winder, editors, *People and Computers IV, Proc. of British SIGHCI*, pages 35–61. Cambridge University Press, 1988.

[10] P. R. Keller and M. M. Keller. *Visual Cues: Practical Data Visualization*. IEEE Computer Society Press and IEEE Press, 1993.

[11] S. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226, 1989.

[12] W. Levie. Research on pictures: a guide to the literature. In D. Willows and H. Houghton, editors, *The Psychology of Illustration: Basic Research*, volume 1, chapter 2, pages 1–50. Springer-Verlag, New York, 1987.

[13] J. Levin, G. Anglin, and R. Carney. On empirically validating functions of pictures in prose. In D. Willows and H. Houghton, editors, *The Psychology of Illustration: Basic Research*, volume 1, chapter 2, pages 51–86. Springer-Verlag, New York, 1987.

[14] G. Lohse, K. Biolsi, and H. Rueter. A classification of visual representations. *Communications of the ACM*, 37(12):36–49, 1994.

[15] W. Mann and S. Thompson. Rhetorical structure theory: Towards a functional theory of text organization. In *TEXT*, volume 2, pages 243–281. Springer, 1988.

[16] J. Marks. A formal specification scheme for network diagrams that facilitates automated design. *J. of Visual Languages and Computing*, 2(4):395–414, 1991.

[17] M. Maybury. Planning multimedia explanations using communicative acts. In M. Maybury, editor, *Intelligent Multimedia Interfaces*, chapter 2, pages 60–74. AAAI Press/The MIT Press, Menlo Park, CA, 1993.

[18] K. McKeown. *Text Generation*. Cambridge University Press, London, 1985.

[19] K. Mullet and D. Sano. *Designing Visual Interfaces*. SunSoft Press, Mountain View, CA, 1995.

[20] S. Roth and W. Hefley. Intelligent multimedia presentation systems: Research and principles. In M. Maybury, editor, *Intelligent Multimedia Interfaces*, chapter 1, pages 13–58. AAAI Press/The MIT Press, Menlo Park, CA, 1993.

[21] S. F. Roth and J. Mattis. Automating the presentation of information. In *Proc. IEEE Conf. on AI Applications*, pages 90–97, 1991.

[22] C. Schmid. *Handbook of graphic presentation*. The Ronald Press, New York, 1954.

[23] D. Seligmann. *A Visual Language for Automated 3D Graphics Generation*. PhD thesis, Columbia University, New York, NY, 1993.

[24] A. Sutcliffe and J. Darzentas. Use of visual media in explanation. In M. Tauber, D. Mahling, and F. Arefi, editors, *Cognitive Aspects of Visual Languages and Visual Interfaces*, chapter 2, pages 105–132. Elsevier Science B.V., Amsterdam, 1994.

[25] A. Treisman. Perceptual grouping and attention in visual search for features and for objects. *J. of Experimental Psychology: Human Perception and Performance*, 8(2):194–214, 1982.

[26] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 1983.

[27] E. R. Tufte. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.

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

[29] R. Wehrend and C. Lewis. A problem-oriented classification of visualization techniques. In *Proceedings of the First IEEE Conference on Visualization: Visualization 90*, pages 139–143. IEEE, Los Alamitos, CA, October 1990.

[30] W. Winn and W. Holliday. Design principles for diagrams and charts. In D. Jonassen, editor, *The Technology of Text*, volume 1, pages 277–299. Educational Technology Publications, Englewood Cliffs, NJ, 1982.

[31] H. Zettl. *Sight Sound Motion: Applied Media Aesthetics*. Wadsworth Publishing Company, Belmont, CA, second edition, 1990.

[32] M. Zhou and S. Feiner. Top-down hierarchical planning of coherent visual discourse. In *Proc. IUI'97*, pages 129–136, Orlando, FL, January 1997.