Perception of pictures without graphical interface

Zdenek Mikovec, Pavel Slavik

Czech Technical University, Dept. of Computer Science and Engineering
Karlovo namesti 13, 121 35 Praha 2, Czech Republic
xmikovec@fel.cvut.cz, slavik@cs.felk.cvut.cz

Abstract. This paper deals with a problem of handling of graphical information in an environment with some limitation. The problem arises when the work with graphical information is limited or impossible at all. The limitations can be either on user side (disabilities) or on the side of a system (small screen in case of handheld devices). This means that it is necessary to apply some effective access to graphical information in such an environment. Textual information can be accessed without any problem and the solution is to work with textual description of the picture. Such a description should allow information filtering, as the structure of graphical information can be complicated. In order to describe a picture in textual form we will use a XML based grammar. Its use will allow us to automize creation of tools for creation of picture description and browsing. A system has been implemented that allows the user to access graphical information without graphical interaction. In the case of handheld devices the textual communication could serve as a preview in order to get closer to the area of interest in a picture. This area can be displayed by limited graphical means. The system has been successfully tested with a good response from potential users.

Keywords: picture description, mobile computing, grammar, XML, graphical interaction, blind users

1 Introduction

The use of pictures in information systems (IS) intensifies communication between the user and computer systems. Due to large flexibility of graphical information it is possible to use it in many applications. On the other hand there are situations when the work with graphical information is limited either by users (disabilities of various kinds) or by a system (small screen in case of handheld devices). In this cases it is necessary to solve effective access to graphical information using textual information without graphical interface. Our solution is based on working with textual description of the picture. Graphical information can have complicated structure and thus we have to be able to filter out unnecessary information using special information filters [9]. Common methods for describing pictures (e.g. alt-text, "D-Tags") are not suitable for description of complicated graphical information, because there is no way how efficiently browse in such a description (specially defining level of detail, filtering information). Our approach is trying to enable perception of more complex pictures. To automize the implementation of a system for manipulation with textual information we use a formal language for description of the structure of the picture description. XML [6] based grammar [1],[7],[8] is used as a tool for formal language definition. The use of XML as an international standard allows us to implement system that can be used in any environment that uses XML.

2 Creation and browsing picture description

The main goal of the specification of methodology of creating picture description is to keep complexity and specific structure of graphical information and allow later efficient browsing. Therefore we choose the object oriented approach. For definition of picture description we used XML [6] based grammar (see section 2.1.4).

Browsing in a picture description is in this work specified as a browsing in a tree (where nodes represent objects and edges represent relations among objects) and filtering out unnecessary information. The user browses in a picture by means of a statement in a special query language based on XML.

2.1 Creating a picture description

The meaning of a picture is specified both by the list of objects in the picture and by relations among these objects. These relations are either of a structural nature (sub-object, left, right etc.) or semantic nature (a person talks to another person). A picture description is then derived from a combination of both relation types.

Graphical information we would like to process consists of two parts:

- objects, their characteristic and behavior
- relations among objects
  - structural (e.g. one object is a part of another one, position, shape)
  - semantic (e.g. behavior, semantic relation).

The formalism used is suitable for pictures with good internal structure that could be expressed in formal way (mostly by means of a graph). The system is oriented for the use in web environment and therefore we will deal in further with pixel oriented pictures.

2.1.1 Objects, their characteristics and behavior

The problem of describing an object is that the description creator can describe the object in different ways. To eliminate differences in object description and speed up describing we have defined a standard for object behavior - basic categories of information that describe objects and their behavior. The browsing user then chooses what characteristics of objects he wants to get (e.g. position, shape, etc.).

There are two basic categories of object description:

- geometrical (position, shape, ...), material, color, weight, category, detail
- action (Behavior of object - "Boy is running", etc.)

2.1.2 Relations among objects

The relations among objects are standardized too:

- semantic - defines semantic relation (e.g. object "foyer" is connected with object "bedroom")
- hierarchical - defines hierarchical relation (e.g. object "bad" is in object "bedroom")
- group - defines group of objects without hierarchical relations

This basic set of categories covers the main descriptions and relations of objects. When describing some special picture there could arise the necessity to define new categories. The possibility to create new categories is incorporated in a design of the grammar of picture descriptions (see section 2.1.4).

2.1.3 Structural and semantic view

When analyzing creation and browsing of the picture description we found out that we could perceive the picture from different points of view, which means that final description could consist of different objects, different object descriptions and different relations. Two types of view were defined:

A. Structural view (see Figure 2) - focused on structural relations

- The goal of this view is to give the user a standard for picture description.
- Major relation is "is in" relation (full arrowhead lines). The objects are ordered into a hierarchical tree. The hierarchy means that the child object is a part of the parent object (e.g. object "bed" is sub-object of object "bedroom" - see Figure 2)
- There are only few of other relations.

B. Semantic view (see Figure 3) - focused on semantic relations

- The goal of this view is to describe significant objects and relations only.
- It is not necessary to order objects into a hierarchical tree as in the structural view.
- This view is useful if the picture should be described from several totally different views (these views consist of different objects and relations). In our example (Figure 1) the semantic view (Figure 3) describes how the rooms of a flat are connected with each other.

2.1.4 Grammar of picture description

A suitable means for handling a complex structure of a picture is the use of a grammar [7],[8]. We will use the formal grammar in the sense of theory of formal languages where grammar is defined as a quadruple G=(N,T,R,S) [1], [8].

Our grammar is developed with the idea of generating a universal description language for representing any objects, their behavior and relations:

```
N = {picture, view, object, description}
T ... alphanumeric symbols
S = {picture}
```

```
R: picture ::= description+ view+
   view ::= description+ object+
   object ::= description+
   description ::= description*
```

```
(Symbol '+' means that the element could be repeated one or several times. Symbol '*' means
that the element could be repeated zero or more times.)
```

`Picture` defines the name of picture. `View` defines the type of view. `Object` defines each object in the picture. `Description` describes object characteristics, behavior and relations among the object. All elements have similar attributes.

2.2 Browsing in a picture description

Browsing in a picture consists of two activities: *browsing* in a tree of objects and *filtering* information. Both activities are described by the user by means of statements in XML query language. Special information filters are used to filter out unnecessary information [9] in order to speed up a process of retrieval of required information. These filters can filter objects, object characteristics and relations among objects.

2.2.1 Information filtering

Users communicate with IS in three steps:

- information type selection (structural or semantic) - see section 2.1.3
- reduction of information - filter out inessential information (objects, descriptions of objects) by defining structural and semantic constraints
- browsing in reduced information using queries

When browsing a picture it is possible to switch dynamically between structural (see Figure 2) and semantic (see Figure 3) view of picture.

Object and relation filter

First method how to filter out unnecessary information is to choose the type of object description we want to see. While there are defined categories of object descriptions:

- geometrical (position, shape)
- material, color, weight
- category, action (connected with), etc.

the browsing user can choose which description categories he wants to see (e.g. color of objects). Applying such a filter will cut out descriptions and objects that do not match the chosen categories.

Second method how to filter out unnecessary information is to define relations between objects we want to see (see section 3). These filters can be combined in a logical expression (using logical operations AND, OR) or applied in a sequence. These filters defined by user are dynamically incorporated into the browser and immediately applied to the browsed description.

3 Example and testing

The example shown below could be considered as an example from the area of mobile computing. As a testing picture we will use scheme of flat (see Figure 1). At first we will analyze created picture descriptions (see Figure 2,3).

3.1 Creating description

We have constructed two types of description of Scheme of Flat (see Figure 1). The first one is structural description (see Figure 2). On the Figure 2 we can see two hierarchical levels of objects. On the first level there are the rooms the flat consist of and on the second level there are objets placed in each room. In general we can define more objects and more hierarchical levels, but for the sake of simplicity we will use two levels only.

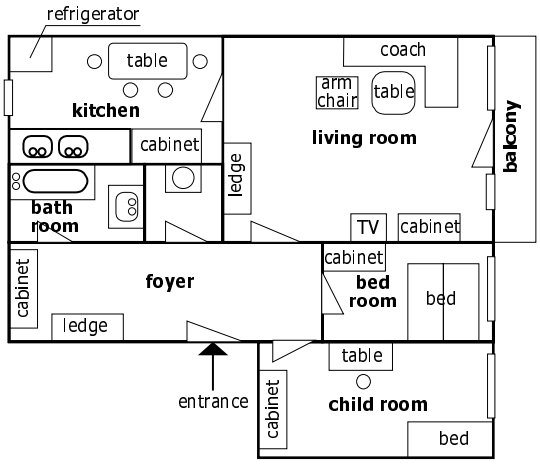

Fig. 1. Scheme of flat

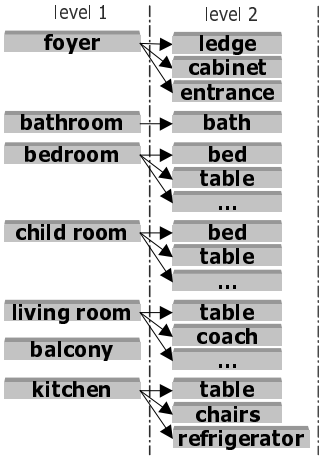

Fig. 2. Structural view

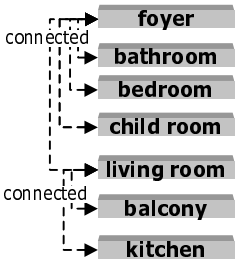

Fig. 3. Semantic view

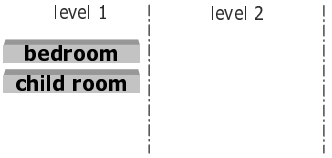

Fig. 4. Filtered objects - structural view

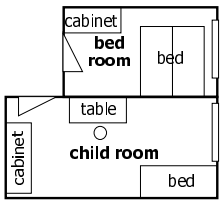

Fig. 5. Filtered objects - graphical view

The second description is called semantic one (see Figure 3). This description defines connections between rooms and will be used for finding way from one place to another one (see section 3.2). Similar relations can be defined between objects in the rooms. They can for example express objects of one suite (e.g. armchair, coach and table from the living room).

3.2 Browsing in description

For representation of objects and relations we use the following notation:

```
Object in the picture: living room
Filtered object: TABLE
Semantic relation: [connected to BEDROOM]
Hierarchical relation: living room
                                table
```

The following examples show possible tasks (a task that consists of a sequence of queries) when navigating in the flat. The queries have a form of statement in a special query language (e.g. XQL, XML-QL). These statements operate over the data structure of a picture described by means of a description language based on XML (see section 2.l.4). The statements in the following examples have only schematic form, because of the complicated structure of the statements of XML query language.

T*as***k** *A***:** *Find rooms wit***h** *windows and bed.*

1. Switch to structural view (Figure 2).
2. Set filter "window" AND "level 1"
   ("level 1" means: Show only objects on the first hierarchical level.)
   Result (sub-graph of graph in Figure 2):
   ```
   BEDROOM
   CHILD ROOM
   LIVING ROOM
   KITCHEN
   ```
3. Set filter "bed" AND "level 1"
   Result (sub-graph of graph in Figure 2) - see Figure 4:
   ```
   BEDROOM
   CHILD ROOM
   ```
4. Show filtered objects ("bedroom", "child room") - see Figure 5
   The information obtained in "textual mode" can be displayed on a small screen (as we found a graphic area of interest).

T*as***k** *B***: H***ow can* **I** *go from entrance to refrigerator.*

We are locating entrance and refrigerator.

1. Switch to structural view (Figure 2).
2. Set filter "entrance" OR "refrigerator".
   Result:
   ```
   foyer
      ENTRANCE
   kitchen
      REFRIGERATOR
   ```

We are finding the way from "foyer" to "kitchen".

3. Switch to semantic view (Figure 3).
4. Set filter "foyer"-"connected to"- "kitchen". (Is there a direct connection?)
   Result: Nothing filtered. => There is no direct connection.
5. Set filter "foyer"-"connected to"-"any object"
   Result:
   ```
   BATHROOM [connected to FOYER]
   BEDROOM [connected to FOYER]
   CHILD ROOM [connected to FOYER]
   LIVING ROOM [connected to FOYER] [connected to KITCHEN]
   ```
6. Set filter "filtered objects"-connected to"-"kitchen"
   Result:
   ```
   LIVING ROOM
   ```

The resultant path is: foyer -> living room -> kitchen

3.3 **I***m***p***lementation*

The system consists of two modules:

- Description creator (module for picture description creation)
- Description browser (module for browsing in picture description without graphical interface)

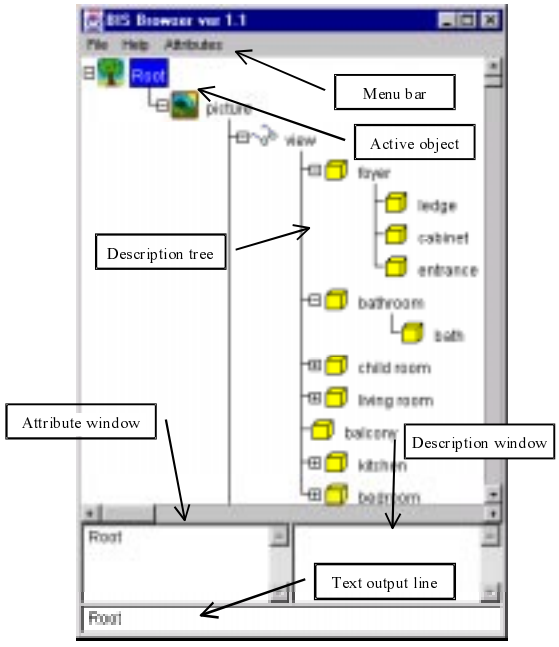

Fig. **6**. Description browser User Interface

For detailed information about implementation of the system see [3],[4],[5].

The system has been implemented in JAVA [2] programming language and the data structure is defined using language based on XML [6]. This means that the system is platform independent (JAVA) and could be used in any environment that uses XML. The description is stored as a part of picture file (e.g. GIF, PNG). On the Figure 6 we can see user interface of Description browser.

The implementation has been intensively tested with group of blind and non-blind users. Both groups of users were asked to get some kind of graphical information when working with textual interface. The tests have shown that the perception of a graphical information without graphical interface does not caused any substantial problems to the users.

4 *Conclusion and future work*

We have defined methodology and also have implemented modules for the picture description creation and browsing the description. While this description contains structural and semantic information we can use special structural and semantic filters [9]. These filters allow us to automate "intelligent" searching in pictures. The use of XML [6] as the definition language for description file structure simplifies the implementation of our methodology in any application.

The future work will deal with testing implemented modules on the platform of mobile computing. We are going to analyze special semantic relations between objects such as relative position of objects (object A is on the left side of object B, object in the neighborhood of object A). The system could be also adapted for the use of vector oriented pictures, where the methods for automatic description generation could be developed.

5 *References*

1. Melichar B.: Languages and Compilers (in Czech), 1996, CTU Prague
2. Chan Mark C., Griffith Steven W., Iasi Anthony F.: 1001 "Java Programmer's Tips", 1997, UNIS Publishing
3. Mikovec Z.: BSc thesis - Picture Description for Blind - Object Library, 1998, CTU Prague
4. Klima M.: BSc thesis - Picture browser for Blind, 1998, CTU Prague
5. Pavlica D.: BSc thesis - Creating Picture Description for Blind, 1998, CTU Prague
6. SGML/XML: http:// wwww.w3.org/XML/
7. Slavik P.: Syntactic Methods in Computer Graphics. Eurographics 83, North Holland, 1983, pp. 134-144
8. Slavik P.: Grammars and Rewriting Systems as Models for Graphical User Interfaces. In: Cognitive Systems. No.4-3, 1997, pp.381-399. ISSN0256-663X
9. Mikovec Z., Klima M., Slavik P.: Structural and semantic dialogue filters. In: Proceedings of the 2nd International Workshop Text, Speech and Dialog, Plzen, Czech Republic, September 1999, pp. 280-285