World Research Journal of Pattern Recognition
ISSN: 2278-8557 & E-ISSN: 2278-8565, Volume 1, Issue 1, 2012, pp.-05-08.
Available online at http://www.bioinfo.in/contents.php?id=184

FUSION METHODOLOGY TO SONIFICATION OF COLOR IMAGES

JAISWAL S.R.*, SHETE N.C., GULHANE P.D. AND GHYAR S.J.

Department of Computer Science & Engineering, JDIET, Yavatmal, MS, India.
*Corresponding Author: Email- im.shraddha34@gmail.com

Received: March 03, 2012; Accepted: May 08, 2012

Abstract- This paper is use for accessing of graphics for visually impaired people. It presents a novel method of sonification of complex graphical objects, such as color photographs, based on a fusion methodology combining sound and speech communication. This approach is supported by a special color model, called semantic color model, which is introduced in this paper. The semantic color model possesses suitable properties that can be used to send the related graphical information in sound or speech in a convenient form. The integration of this approach with the annotated SVG format developed within the ongoing GATE project, which is also briefly described in paper, enhances the effectiveness of the system.

Keywords- Fusion methodology, Color Models, Sound Interpretation, Speech Interpretation, Image Sonification, Hybrid Dialogue Processing, Visually Impaired.

Citation: Jaiswal S.R., et al. (2012) Fusion Methodology to Sonification of Color Images. World Research Journal of Pattern Recognition, ISSN: 2278-8557 & E-ISSN: 2278-8565, Volume 1, Issue 1, pp.-05-08.

Introduction

One of the major problems for visually impaired people is Accessibility of graphics. Recent research for this is concerned with both to the use of tactile devices [2] and developing methods for sonification of various types of graphical objects [1].

Even supposing the use of tactile devices has some advantages and can be better for some special tasks, this approach seems to be limited and is not directly applicable for more difficult graphical objects such as photographs.

The methods for sonification can be divided into two groups. The first group arranges the graphical data defining a graphical object into a predefined order in which they are converted into sound. The second group, query based methods, utilizes a user controlled pointer to select a region of the studied image. The system then assigns an appropriate sound to this selected region. Examples are Helle's system representing mobile phone menus, sonification of heating schematics and the Audio graph system for exploring graphs using a touch panel.

In our paper, we introduce a semantic color model, which forms a framework for a novel approach to sonification, which is applicable to both vector and raster images. A classic image that we have in mind is a color photograph. We assume that the image is represented in the annotated SVG format, in the context of the ongoing project GATE. This is however no limitation, because every standard graphical format can be converted into SVG format, which is a special case of the annotated SVG format. The approach is in fact not dependent on the assumed graphical format. Nevertheless, thinking about sonification in the context of the annotated SVG format is relevant and advantageous, which is why the next Section briefly describes the project GATE and the annotated SVG format.

Gate Framework

GATE (Graphics Accessible To Everyone) is an ongoing project currently being developed at the Faculty of Informatics, Masaryk University Brno, Czech Republic. This project is pointing to enable handling computer graphics by means of dialogue. It covers the development of advantages arranged for easy picture explanation, provision of blind users with support for viewing pictures, and finally, the development of a system developed for generating images by means of a dialogue and supporting the blind to create some limited form of computer graphics.

The GATE project develops two technologies to handle explained graphics. The SVG format stores graphical content, either vector or raster, as well as the explanation information, whereas graphical ontology represents a knowledge base, which contains the interpretation of the explanation. A description of graphical ontology and its coupling with SVG can be found in [9]. SVG and Ontology modules provide important edges for the manipulation with the graphical and semantic information of a picture.

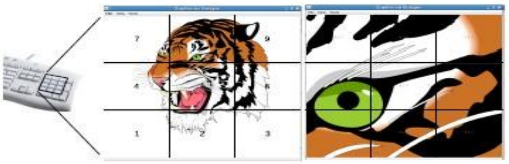

Fig. 1- Recursive Navigation Grid

Recursive Navigation Grid [7], RNG, represents the navigation stamina of the system. The image is divided into nine identical rectangular sectors analogously to the numerical keys 1-9 of the numerical keyboard. Each sector can be subdivided in the same way recursively. The RNG module enables the user to carry out ”zooming” by traversing the grid (see Figure 1) and thus obtain the applicable pieces of information with the accuracy demanded. GATE supports two basic approaches of recovering information and informing the user about the graphical content in a non-visual way. These strategies are of two separate modules, GUIDE and EXPLORER.

The module GUIDE is to provide verbal information about the picture, using both the pieces of information obtained by tagging the picture and the pieces of information gained directly from the picture format. The GUIDE module cooperates with RNG in order to recover the current navigation state. On lower levels of the navigation grid, the GUIDE module provides more details than on upper levels. Verbal Information Module (VIM) is responsible for the verbal part of the dialogue. It is developed by the GUIDE module and solves possible misunderstandings in the communication. VIM is also able to process special commands related to What-Where Language (WWL). WWL is a simple fragment of English. Each sentence of this language has the form of WHAT is WHERE or WHERE is WHAT. It allows the user to ask simple questions about the objects in the scene or image and their position, e.g. ”What is in the central part?”, ”Where is the flower?”, ”What is in the background?”, etc.

Unlike GUIDE, the communication of EXPLORER is not primarily verbal, but analogue. It is controlled by means of a mouse, digitizer, or numerical keyboard. The output sound information is also primarily non-verbal. The nonverbal communication allows the user to perform a quick energetic study of the non-annotated details of the image. Methods of conversion of colors into sounds is discussed in details in the second part of the papers. The GATE system is structured in accordance with the requirements of the client-server architecture. As an alternative of developing big separate applications, the GATE system supports the application logic to be run on a remote application server.

Speech and Sound Processing of Graphics

For study process of pictures firstly, the user is provided with brief essential information about the picture. The technical information, like the size in pixels, compression type, etc., are obtained automatically from the data stored in the format of the picture. Then the user starts the exploration by using the picture information tree structure, into which the data describing the pictures are organized, navigation by keywords, querying by means of WWL, and the use of the EXPLORER module. In the following

example a blind user explores a color photo (photo of Jeseniky Mountains - see Figure 2). The characters in the pictures stand for: A: a cloudy sky, B: snow-covered landscape, C: the Praded Mountain, D: TV transmitter Praded, A1: bright cloud, A2: dark clouds, B1, B2 and B3: snowy trees, B4: snowy forest, B5: the Svycarna chalet,

X: foreground, Y: middle part and Z:background.

U: (Enters the system by choosing the picture.)

S: This is a color photo of Jeseniky Mountains in winter nearby the Svycarna chalet. For more information, choose Description, Technical information, use keywords or What-Where language.

U: Technical information.

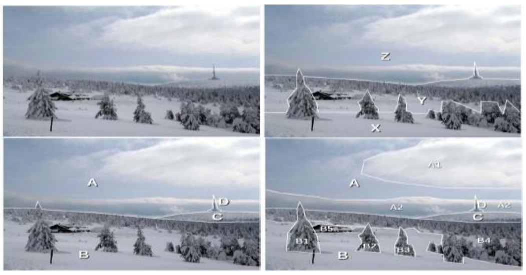

Fig. 2- The original photo (upper-left), the tagged depth (upper-right) and the tagged first (lower left) and second (lower-right) semantic levels

S: This is a color photograph in full color. The size is two megapixels....

U: What is in the central part?/Comment: The user asks a typical WWL question. The”Y” area and the first semantic level are taken into consideration./

S: In the central part, there is a snow-covered landscape, the Praded Mountain and the Praded TV transmitter.

U: What is in the landscape? /Comment: The second semantic level is taken into consideration./

S: There are three snowy trees in the foreground, snowy forest in the central part and right-hand foreground and the Svycarna chalet in the central part.

U: (Because there are no more semantic details about the central part, the user starts the direct exploration of the picture by means of the EXPLORER utility. This module provides the information about the investigated graphics by means of sonification and auralization based on the principles described in what follows).

Sonification

Sonification can be roughly divided into two types of sonification. In high-level (symbolic) sonification, visual information is translated into natural language. In contrast, low-level sonification transposes visual information into an abstract audio signal.

High-level Sonification

The majority of work on sonification for the blind has focused on high-level (symbolic) sonification. Text -to-speech (TTS) is the most well known sonification system, where such software as the VoiceOver function on Apple products and JAWS (Freedom Scientific, Inc.) can sonify text characters and objects displayed by a computer. The advantage of such systems is that they map visual information to the information-rich realm of the natural language. The obvious limitation of high-level sonification is that it is limited to objects that have obvious semantic representations. For example, it is not clear how to sonify complex shapes, color variations and detailed textures.

Low-level Sonification

While high-level sonification eases the burden of recognition, it is also limited by the lexicon of the system. In contrast, a mapping to an abstract audio space (low-level sonification) has the advantage of dealing with a wider range of objects without being constrained by a lexicon. Low-level sonification can still work with hard-to-label (untrained) objects and can work in realtime without relying on remote guides.

The vOICe system [13] sonifies the global luminance of an image and maps luminance values to a mixture of frequency oscillators. Specifically, the image brightness is mapped to amplitude and location is mapped to a frequency. The vOICe system scans the entire field of view of a head mounted camera with a vertical bar from left to right and transposes the luminance over the vertical bar to sound. One of the advantages of the vOICe system is that it sonifies an entire image to convey the global content of any type of scene and the system does not require any type of prior training or lexicon.

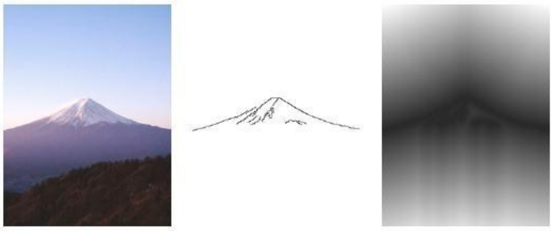

Fig. 3- (a) source image (b) edge image (c) distance image

Color Models-Assigning Sounds to Colors

For the blind person, sound is a natural way for taking the information of colors. Both non-verbal sound and speech can be used and combined. To assign sounds to colors, it is natural and practically expected to choose a color model as a basis for further consideration. Let us briefly discuss basic possibilities.

RGB (for Red, Green, Blue) Color Model

To understand the RGB model, the blind users have to know that each color in the color space is uniquely defined by its red, green and blue component and what colors we get by mixing RGB components. This model is relatively easily to understand.

HSL AND HSV Color Models

HSL (for Hue, Saturation, Lightness) and HSV (Hue, Saturation, Value) color models are mapping colors as points in a cylinder. It is more accurately than RGB and are popular and, their description is less natural in comparison to RGB model. Moreover, the conversion from RGB to HSL or HSV and vice versa is given by the transformations that are not continuous.

Semantic Color Model

We previously discuss for standard color models which shows that the most suitable standard color model is the RGB model, That can be used as transformation of colors into sounds. The main drawback of using a trichromatic color model is that each color is described by three primary colors. In some cases, for instance if we think about brown color, to identify the color we must know the corresponding proportions of the primary colors and correctly estimate the proportions from the combination of the corresponding sounds. This may be uneasy and lead to misunderstandings. However, if we take a color that belongs to the primary colors, there is no such problem.

This inspires us to consider creating a color model with more primary colors, enabling a reasonable set of colors to be described using just two primary colors.

Primary colors for the semantic color model are: red, green, blue, yellow, orange, purple, brown and gray. The semantic color model uses the RGB model as its basis. Concrete RBG components (X_{rgb}, Y_{rgb}, Z_{rgb}) are assigned to primary colors of the semantic color model in such a way that $X_{rgb} + Y_{rgb} + Z_{rgb} = 1$. Let $P = \{R, G, B, Y, O, P, BR, GR\}$ be the set of the corresponding three-dimensional vectors. The idea is to uniquely express a color by means of only two primary colors. It is easy to distinguish two sounds and describe approximately the color that is expressed in this way. Formally, the transformation of RGB model into semantic color model is given as follows:

For the color C having the RGB components (αX_{rgb}, αY_{rgb}, αZ_{rgb}) satisfying $X_{rgb} + Y_{rgb} + Z_{rgb} = 1$, find two vectors from the set P that have minimal Euclidian distance from the vector $U = (X_{rgb}, Y_{rgb}, Z_{rgb})$ (if there are more such vectors, choose the first two in alphabetic order of the set P; this situation, however, has very low probability).

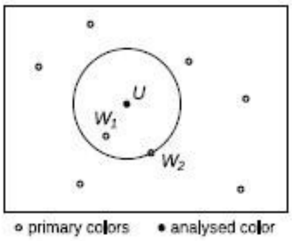

Fig. 4- Schematic two-dimensional projection of the semantic color model

Denote these vectors W1 and W2. Then Compute the Euclidian distances $d(U,W_1)$ and $d(U,W_2)$ of the vector U and vectors W1, W2. Let W1 and W2 correspond to the color with i-th position and j -th position in the set P respectively and let $(X_1, \ldots, X_{|P|})$ denote the coordinates of the vector U in the semantic color model.

Then put

$X_i = d(U, W_1) / (d(U, W_1) + d(U, W_2))$,

$X_j = d(U, W_2) / (d(U, W_1) + d(U, W_2))$

and $X_k = 0$ for k is not = i ∧ k is not= j.

The values αX_i and αX_j are the color components of the color C in the semantic color model. Note, that each RGB vector can be uniquely expressed in the form $(\alpha Xrgb, \alpha Yrgb, \alpha Zrgb)$ where $Xrgb + Yrgb + Zrgb = 1$ and α is a non-negative real number. The color space of the semantic color model is a proper subspace of the RGB color space and there is no injective transformation from the color space of the semantic color model to the RGB color space.

Semantic Color Models and The Description and Investigation of Colors

Each color of the color space of the semantic color model can be uniquely described by the components of exactly two colors belonging to the set of primary colors. There are two ways how to exploit this property. First, by sound interpretation of colors, and second, by speech interpretation.

Sound Interpretation of Colors in the Semantic Color Model

In this mode, the blind user can browse freely within a graphic by means of a digitizer, keyboard (using the arrow keys of the numerical panel for navigation), mouse or touch screen, and is provided by an analogy sound representing the color of the selected region. This region can be either the pointed pixel, or the pointed pixel and its selected neighborhood. In the second case, the user defines the region by telling the system the maximal distance between the pointed pixel and any pixel belonging to the region. The color of the region is taken as the mean color computed from all the pixels belonging to the region.

The boundary of the image is signalized by a special sound. For the explained SVG format, the user's direction is supported by the explanation, provided by the format.

The semantic color model provides the user with the information about the approximation of given color by means of two closest primary colors. Specific sounds are assigned to each of the primary colors. Intensity of the sound expresses the intensity of the corresponding parameter.

The sounds assigned to the primary colors can be artificially generated sound, sounds of musical instruments, or other sounds that can be equally well distinguished. Besides the fundamental frequency of the sound, timbre (tone quality) can be used to generate the set of sounds representing primary colors. The following attributes determine the timbre of the generated sound:

1. the range between tonaz and noise-like character,
2. The spectral envelope,
3. the time envelope in terms of rise, duration, and decay,
4. the changes both of spectral envelope and fundamental frequency,
5. the prefix, an onset of a sound quite dissimilar to the ensuing lasting vibration.

Adding a new color to the set of primary colors is an option. In this case, the user selects the color they want to add to the set of primary colors and assigns a sound to it. This can be useful when the user wants to find the same or similar color in the picture. In the semantic color model, adding or removing a new color to the primary colors is a simple and intuitive operation.

Speech Interpretation of Colors in the Semantic Color Model

Sonification of colors enables fast browsing and investigating the analyzed graphic. On the other hand, semantic color model is suitable also for providing the verbal information about the colors. Generally, this information, as obtained directly from the semantic color model, has the form color1 = x; color2 = y, where x and y mean the intensity of the first primary color and second primary color, respectively. Note that in this mode the colors, which are added as new primary colors to the set of primary colors by the user and to which no names are assigned, are ignored. The main drawback of this mode is, of course, that delivering the verbal information in speech form is very slow. On the other hand, the information is precise and explicit, and need not be interpreted by means of an analogy.

Speech interpretation of colors is a corresponding tool to the sound interpretation of colors. Enabling the user to mix freely both strategies gives them, together with the explained structure of the graphic in the explained SVG format, a flexible tool for analyzing the image.

Conclusion

The semantic color model provides a suitable framework for sonification of complex graphical objects. The preliminary testing shows that the approach is promising, much more testing has to be done especially in connection with the problem of choosing appropriate sound encoding the primary colors and lightness. Further, it is clear that the meanings and comments of the visually impaired users are of the maximum importance.

References

[1] Bulatov V. and Gardner J. (2004) *In 3rd Annual Conference on Scalable Vector Graphics*.

[2] Daunys G. and Lauruska V. *Int. Conf. ICCHP*, 12-15. Berlin.

[3] Fredj Z.B. and Duce D. (2003) *Theory and Practice of Computer Graphics*, *IEEE,* 49-55.

[4] Glassner A. (1995) *Principles of Digital Image Synthesis,* I and II.

[5] Jones W.D. (2004) *IEEE Spectrum*, 11-12.

[6] Kaczmarek K. (2004) *Research Report* 5-R01-EY10019-08, University of Wisconsin.

[7] Kamel H.M. and Landay J.A. *Proceedings of the fifth international ACM conference on Assistive technologies*, 33-40.

[8] Kennel A.R. (1996) *Proceedings of ASSETS*, 51-56.

[9] Kopeˇcek I. and Oˇslejˇsek R. (2006) *Proceedings of the Second IASTED Int. Conf. on Computational Intelligence*, 343-348.

[10] Kopeˇcek I. and Oˇslejˇsek R. *Int. Conf. ICCHP*, 295-302.

[11] Matta S., Rudolph H. and Kumar D.K. *13th European Signal Processing Conf. Antalya*.

[12] Satoshi I. (1996) *ACM SIGCAPH Newsletter* 55, 16-21.

[13] Meijer P.B.L. (1993) *IEEE Transactions on Biomedical Engineering*.