Visual Storytelling Warehouse

Yutu Liu
Computer Science Department
Texas A&M University
alinux@tamu.edu

Ergun Akleman
Visualization Sciences Architecture Department
Texas A&M University
ergun@viz.tamu.edu

Jianer Chen
Computer Science Department
Texas A&M University
chen@cs.tamu.edu

ABSTRACT

In this paper, we introduce the concept of Visual Storytelling Warehouse, an interactive XML-based automatic cartoon-style visual storytelling system to generate new stories based on written story structures and user's requirements. The Warehouse will consist of the visual and no-visual elements including characters, events, background images, and story plots. Each of these elements will be associated with some semantic meanings that can be easily interpreted for constructing stories by Visual Storytelling Warehouse. Stories will be written based on a set of XML schemas which specify essential elements to construct visual stories. Given an abstracted story structure, the Visual Storytelling Warehouse will be capable of producing multiple versions of visual stories by taking characters, background information and story plots from its database based on defined specification from the users. Each of those stories might have a variety of story elements such as different characters, events or backgrounds. The proposed Warehouse will be a highly adaptive system that can compose nested visual stories from its existing story plots. Therefore, complicated visual stories with multiple events will be generated on-the-fly by the system using simple story plots. Eventually, the Warehouse will evolve into an open system with standards that can allow artists and writers to share cartoons and stories for building a more dynamic community where more comprehensive visual stories can be produced through WWW around the world. The standards will include a set of XML schemas defining character, events, backgrounds and story structures etc.

Categories and Subject Descriptors

I.3.2 [**Graphics Systems**]: Stand-alone systems

General Terms

Theory

Keywords

Multimedia, Visual Storytelling, Cartoon

1. INTRODUCTION

Visual Storytelling is one of those oldest mastered skills by human kind. It has been used to record many events by many ancient tribes [8]. There should be no doubt of the role of visual storytelling in the development of our civilization on this planet. Visual storytelling is still being used everyday by us to tell stories due to its power of expression and the easiness of passing message to its audience. However, with today's technology, can we find new ways to do visual storytelling? More specifically in this paper, is there a more efficient and economic way to create visual stories? Even with the aid of today's powerful computer software, it is still a very time-consuming process to create visual stories. The difficulties of creating visual stories for casual users lie in the following facts such as writing storylines, and the creation of image itself. It will be impossible to expect ordinary users to do the job well without years of training.

Our proposed Visual Storytelling Warehouse will be able to help those un-trained casual users to create their visual stories without even drawing a single line. Of course, they still have to make up the storylines, but more complicated storylines can be built based predefined storylines or story plots in the database of Visual Storytelling Warehouse. One benefit of using the warehouse is that it can generate many different versions of stories based on the same story structure by simply changing characters and backgrounds etc. An interactive storytelling system can be developed to give children or even adults the right tool to create their cartoon-style stories easily. One possible application for the Visual Storytelling Warehouse is providing elementary school teachers with this tool to create interesting visual stories or even by children themselves to help them to learn.

2. RELATED WORK

Many researches have been done on visual or none-visual storytelling. Some of them focus on the techniques of visual storytelling such as lighting and transition etc. [1, 5]. Gershon's research discussed the possibility of using storytelling for the advance of information visualization [7]. Neal talked about issues related to visual storytelling at distance [11]. Some research has shown the benefits of using visual storytelling techniques for teaching algorithms in computer science [9]. Vaucelle and Davenport proposed a framework for digital video visual storytelling. Their frame is a text-driven system that allows user to create movie from its video

database in real-time according to user's text input [17]. It's a very interesting idea, but it's unlikely to happen soon since it will have many obstacles to overcome. To build a visual storytelling system, it needs a large database of images associated with semantic meanings. How to retrieve those semantics from image is not an easy job. Many researchers have tried different methods to extract those content from various types of images [4, 10, 12, 16]. However, it's still a unsolved problem.

3. SYSTEM OVERVIEW

3.1 Architecture

The Visual Storytelling Warehouse consists of a database, an interpreter, an assembler and an interface as shown in Figure 1. The database stores the images for different characters, background images, and predefined story structures. The interpreter's job is trying to match semantic meaning of the presented user's requirements with the associated meaning of elements from the database based on a matching score mechanism. The assembler basically puts all of those matched elements from database together to build a working visual story. The user interface is the place users can interactively put up their requirements and to see their finished stories.

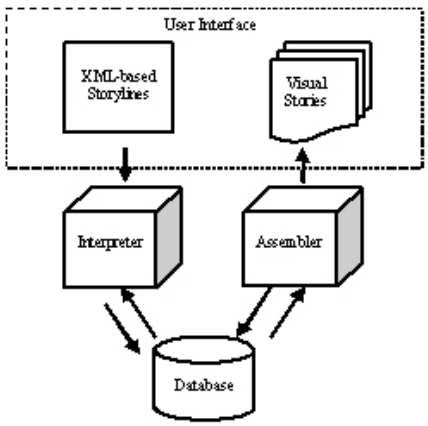

Figure 1: Architecture of Visual Storytelling Warehouse

3.2 Formalization

Let s be a story, and i_j be jth image in the story. Therefore, a story can be represented by a group of orderly images as: $s = i_s > ... > i_e$. Note that i_s and i_e stand for starting and ending image respectively. Given c_j as a character, the set of n characters inside the database will be $C = \{c_1, ..., c_n\}$. Given b_j as a background image, the set of n background images inside the database will be $B = \{b_1, ..., b_n\}$. Given t_j as a text balloon, the set of n balloons inside the database will be $T = \{t_1, ..., t_n\}$. Based on defined notations on character, background image and balloon text, each composited image can be expressed logically as the following:

$$im_j = \{c_h, ..., c_n \in C\} \cup \{b_m \in B\} \cup \{t_k, ..., t_l \in T\}$$

For each element, it's necessary to have a set of attributes such as name, size and mood to describe them. Let $A = \{a_1, ..., a_n\}$ be a set of n attributes, then A_{im_k} stands for the set of attributes of the k th image in the database. Similarly, A_{c_k}, A_{b_k}, A_{t_k} and A_{s_k} are for the attributes set for k th character, background, text balloon and story in the database respectively. Based on the attribute set of each element inside the image, the attribute set of this image can also be expressed logically as the following:

$$A_{im_j} = \{A_{c_h}, ..., A_{c_n}\} \cup \{A_{b_m}\} \cup \{A_{t_k}, ..., A_{t_l}\}$$

Furthermore, a story's attribute set can also be expressed as the following:

$$A_{s_j} = \{A_{im_s}, ..., A_{im_e}\}$$

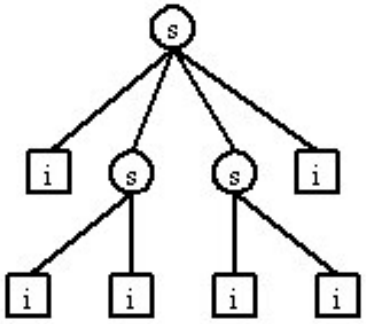

Figure 2: Story Structure in Tree

We have found that story structure can be expressed as a tree that is more intuitive for us to understand and construct more complicated stories. As shown in Figure 2, each circle, the s-node, stands for a story or sub-story and each square, the i-leaf, for a composited image. By representing the story in a tree structure, one can easily see how the stories to develop into more complicated ones. A story can expand with its branches of sub-stories. By tracing the tree depth-first, one can produce a story with a set of composited images that simply consists all branches from the tree in the order of being traced.

4. ESSENTIAL BUILDING BLOCKS

4.1 Character

Characters are the basic elements for building visual stories in the Visual Storytelling Warehouse system. To ensure the quality of the composed stories, it needs to store certain amount of images with various attributes. The majority of those images will be obtained through internet search engines such as Google, and from other large 2D image [3, 2] and 3D model [14, 13] databases. Each image is associated with a set of attributes such as size, mood, direction etc for the interpreter to match semantic meanings between characters and user's requirements.

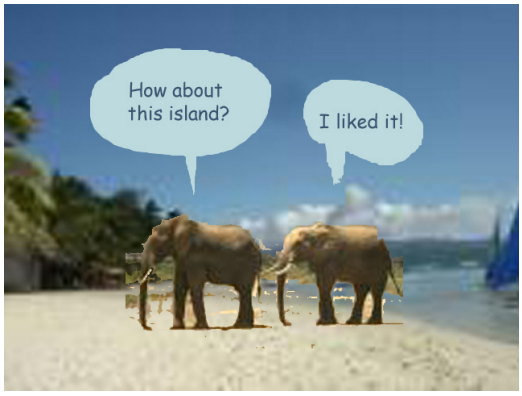

Figure 3: Character Image

During the process of building image database, we will automatically eliminate the background of the characters. Since story does not depend of the quality of images, we do not need to eliminate the background completely. Elimination can be very rough and in fact we can even allow artifacts as shown in the Figure 3. Note that in this image, the artifacts around the legs of elephants are very visible, but it does not hurt the meaning of the story.

4.2 Background

Background is another basic element for visual storytelling. It usually sets the tone of the story and gives its audience enough information to catch the story. The background image will also be grabbed from Google as a low quality image. These images will be scaled up to have an uniform size. By doing it, we will create an extremely low quality image such that we will not have any copyright problem. As seen in the Figure 4, the quality of background image will be very low. Again, a set of attributes associated with each background image will be stored into the database.

Figure 4: Background Image

4.3 Balloon and Text

Balloons, texts and titles will automatically created as images given users' requirements of characters and plots. The color of balloons and colors of texts will be defined by colors of background image. For examples of text balloons, see Figure 3 & 4.

4.4 Composited Image

Composited images are the powerful information-rich building blocks in the system. A composited image will be created by composing characters, text balloons and background as shown in example Figure 3 & 4. They have contained all the information passed from characters and background images and text balloons, and are ready to be put in order to tell the story.

4.5 Story

A story consists of a sequence of images or a mixture of images and predefined stories in an orderly manner. There is a difference between story and story structure. A story is just an instance of a story structure. Story structure is an abstract idea of a story that defines the order of images, and attributes of characters, backgrounds and images etc. For example, a story structure of the "Migrant Elephants" story can be defined as a sequence of composited images and several characters and text balloons associated with those images. Inside the story, several characters arrive at a new location. It causes uneasiness among old residents, then they are accepted by old residents. At the end, something happened. See Figure 5 for the story. Each of those elements will only defined with attributes that allows many of those elements to be changed to fit into different stories. As shown in Figure 6, by simply changing the character to hippo, we have created a new story with the same story structure. The warehouse might be able to create multiple versions of the story, some of those might work and some of them will not make any sense at all. Possible predefined stories can include many classic ones like "escape, rescue, underdog, revenge, love and disaster etc. [15]"

5. ALGORITHM

5.1 Matching Score

Given the user-defined story structure, the interpreter will try to match elements from the warehouse's large database with requirements from the users. Only the top scorers will be returned for the final construction of the visual story. The score will be calculated as the following. Let user defined attribute set be $A_{ur} = \{A_{uc} \cup A_{ub} \cup A_{ut}\}$. Note that A_{ur}, A_{uc}, A_{ub} and A_{ut} are user defined attribute sets for the overall attribute, character, background and text balloon attributes. The matching score will be the maximum of total score between user defined attributes and the ones of characters, backgrounds and text balloons from the database as the following:

$$M(A_{ur}) = max(M(A_{uc}, A_{c_i}) + M(A_{ub}, A_{b_j}) + M(A_{ut}, A_{t_k})).$$

The calculation of $M(A_{uc}, A_{c_i})$ is the following:

$$M(A_{uc}, A_{c_i}) = \sum_{i=1}^{n} M(a_{uc_i}, a_{c_i})$$

Note that the score for individual attribute is the following:

$$M(a_{uc_i}, a_{c_i}) = \begin{cases} 1 & \text{if } a_{uc_i} = a_{c_i} \\ 0 & \text{if } a_{uc_i} \neq a_{c_i} \end{cases}$$

It will be very similar as above for the calculation of $M(A_{ub}, A_{b_j})$ and $M(A_{ut}, A_{t_k})$.

5.2 Complex Story

We have found that stories can be composed based on existed stories and composited images. Based on notations defined in section of 3.2, it will be more intuitive to express the complex story as a tree structure as shown in Figure 2 & 7. We will view any story as a tree structure with ordered branches. As shown in Figure 7, after removing one story node, one can create a new story. Similarly, a story node can be expanded to create a new one as well. In real application, the users will be presented with this story structure, by clicking one of those nodes, users can remove or replace that node with a new one. It looks promising for us on this path to create new stories based predefined stories from database. Take the "Migrant Elephants" story as an example, by simply adding a new story like "Escape" or "Revenge" to the end, it evolves into a total different new story. Complex nested stories might have some side-effects if a branch is replaced by a story node since the logic among composited images might not be consistent here. We might be able to deal this problem with a more comprehensive representation of story structures later.

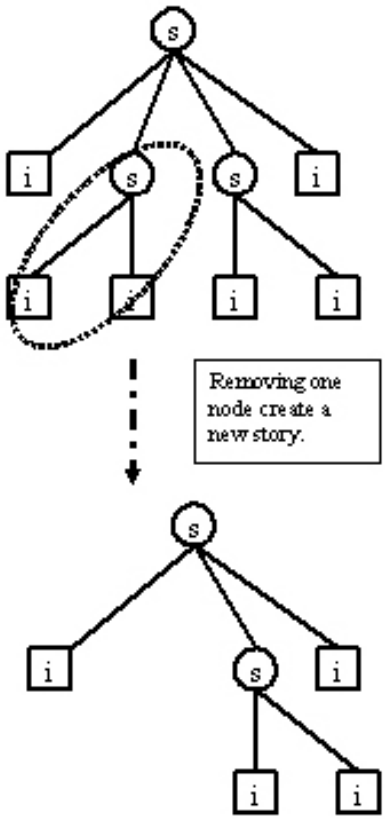

Figure 7: Background Image

6. DISCUSSION

In this paper, we have touched one of those grand challenges in the field of visual storytelling. Although our proposed system is far away from real automatic system, it is close. Unlike many proposed automatic animation systems requiring a stream of high quality video images, our system is more likely to be implemented as a working one since it only needs a sequence of low quality images. Additionally, it has the help of many text balloons to convey the message. However, it still offers us the true challenge to make it work eventually. Clearly, our contribution is presenting the concept of building an efficient and economic method of visual storytelling.

It will not be an easy job to fill the database with images associated with semantic meanings. Many of those semantic meanings need to be extracted from images manually. It's possible to build the image database using an AI system to automatically search images and extract semantic meanings. Many researchers have proposed some methods to search and retrieve semantic meaning of images automatically [4, 10, 12, 16]. It's also possible to build the semantic image database by taking images from some controlled animation models including human and animals [6]. This approach will save a lot space for generating those images in real-time. This will also generate better composited images since it can create character images in a continuous mode carrying subtle meanings. There will be other issues besides technical ones, the copyrights of those images obtained from web might be trouble even they are processed as low quality images. Given the XML-based schemas, it will not be very difficult to extend the Visual Storytelling Warehouse into an open system that allows many different systems share their images and stories to build more interesting stories in real-time. It will be very exciting for us to see this system to be implemented and tell some great stories one day.

7. REFERENCES

[1] G. A. A change of scene. *Computer Graphics and Applications, IEEE.*, 21, May-Jun 2001.

[2] A. D. B. C. Colombo and P. Pala. 3d image databases: Acquisition and retrieval by content. *In Proceedings 4th International Workshop on Multimedia Information Retrieval MIR 2002*, 2002.

[3] R. Campbell and P. Flynn. A www-accessible 3d image and model database for computer vision research. *Empirical Evaluation Methods in Computer Vision*, pages 148–154, 1998.

[4] W. W. Chu, A. F. Cardenas, and R. K. Taira. Knowledge-based image retrieval with spatial and temporal constructs. *International Syposium on Methodologies for Intelligent Systems*, pages 17–34, 1997.

[5] M. EI-Nasr and I. Horswill. Expressive lighting for interactive entertainment. *Multimedia and Expo*, 2003.

[6] D. M. Evangelos Kokkevis and N. I. Badler. Autonomous animation and control of four-legged animals. *Graphics Interface, IEEE.*, 1995.

[7] N. Gershon. What storytelling can do for information visualization. *Communication of the ACM*, 44:31, 2001.

[8] R. R. Hollowayis. *Archaeology of Ancient Sicily*. Routledge, 2000.

[9] C. D. Hundhausen and J. L. Brown. Personalizing and discussing algorithms within cs1 studio experiences:an observational study. *Proceeding of International Computing Education Research Workshop*, 2001.

[10] J. Jeon, V. Lavrenko, and R. Manmatha. Automatic image annotation and retrieval using cross-media relevance models. *In ACM SIGIR 2003.*, 2003.

[11] L. Neal. Storytelling at a distance. *eLearn*, 5, 2001.

[12] A. Oliva, A. Torralba, A. erin Dugu, and e erault. Global semantic classification of scenes using power spectrum templates. *Electronic Workshops in Computing*, 1999.

[13] M. K. Philip Shilane, Patrick Min and T. Funkhouser. The princeton shape benchmark. *Shape Modeling International*, 2004.

[14] A. PROJECT. Aim @ shape. http://www.aimatshape.net/research.

[15] T. B. Ronald. *20 Master Plots and How to Build Them*. Writer's digest books, 1993.

[16] C. Schmid. Constructing models for content-based image retrieval. *In Proc.IEEE Conference on Computer Vision and Pattern Recognition, 2001.*, 2001.

[17] C. Vaucelle and G. Davenport. An open-ended tool to compose movies for cross-cultural digital storytelling: Textable movie. *In Proceedings of ICHIM 04 'Digital Culture & Heritage*, 2004.

Figure 5: This is a story with Elephants and frogs as characters

Figure 6: This is the same story with Hippos and frogs as characters