Making Digital Signage Adaptive through a Genetic Algorithm
Utilizing Viewers' Involuntary Behaviors

Ken Nagao and Issei Fujishiro
Graduate School of Science and Technology, Keio University, Minato, Tokyo, Japan

Keywords: Digital Signage, Image Processing, Genetic Algorithm, Human Behavior Recognition.

Abstract: Digital signage has been becoming more popular due to the recent development of underlying hardware technology and improvement in installing environments. In digital signage, it is important to make the content more attractive to the viewers by evaluating its current attractiveness on the fly, in order to deliver the message from the sender more effectively. Most previous works for this evaluation do not take the viewers' feeling towards the content into account, and the content is improved manually if needed in an off-line manner. In this paper, we present a novel method which does not rely on such manual evaluation and automatically makes the content more adapted to the viewers. To this end, we take advantage of the viewers' involuntary behaviors in front of the digital signage for online updates through the usage of a genetic algorithm.

1 INTRODUCTION

Digital signage has been becoming more popular due to the recent development of underlying hardware technology and improvements in installing environments. Digital signage is an electronic display which shows various messages to the public, and there can be found many advantages to use the digital signage compared to the traditional paper signage. Indeed, digital signage can be controlled in real time, and we can instantly replace the content. As a new interactive display device, digital signage has recently attracted much attention from the field of computer vision.

Digital signage has to provide "attractive" content to the potential viewers in order to deliver the original message effectively. Therefore, it is important to modify the appearance of the content continually. In general, this modification is done by humans because it is necessary to allow for the subjective matters as well as aesthetic issues. We herein propose a method which does not rely on these kinds of manual processes and adaptively improves the content. The method utilizes a genetic algorithm (GA) to learn "what content is attractive" and automatically make the content more attractive to the viewers. In the system, "content designs" are regarded as "individuals" in the GA.

In order to utilize a GA, there must be an evaluation mechanism installed in the system. In digital signage, "evaluation" should be based primarily on the viewers' feeling towards the digital signage. To this end, the proposed method is designed so as to make use of the viewers' involuntary behaviors in front of the digital signage. For example, a viewer may change his head angle or move his eyes for reading each section of the content. If the section was convincing he would nod or if it was unacceptable he would shrug his shoulders. Such involuntary behaviors come directly from his emotion, and hence, if the digital signage system can capture and recognize these behaviors of the viewer, it can automatically learn his actual feeling toward each section of the content. Moreover, as digital signage is for pedestrians, many samples for the evaluation can be collected readily.

In order to empirically evaluate the proposed method, we have developed a pilot system for displaying academic conference posters. This is because messages and relationships of each sections in academic conference posters are clear and easy to deal with. Figure 1 shows an example of adaptive conference poster, and the proposed system in actual use can been seen in Figure 2.

The remainder of this paper is organized as follows. In section 2, we review related work. Section 3 describes the detailed approach adopted in the proposed system. Finally, Section 4 concludes the paper with brief remarks on future work.

2 RELATED WORK

So far, direct questions to viewers, such as interviews or questionnaires, have been commonly used for eval-

Nagao K. and Fujishiro I..
Making Digital Signage Adaptive through a Genetic Algorithm - Utilizing Viewers' Involuntary Behaviors.
DOI: 10.5220/0004346100540059
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2013), pages 54-59
ISBN: 978-989-8565-48-8

uating the content of digital signage (Alt et al., 2012). One major drawback in such direct questions lies in that much human labor is required for gathering and analyzing the data. In addition, the viewers do not always tell their actual feeling with such off-line styles of questions.

There are many known attempts to recognize human faces in video (Pantic and Rothkrantz, 2000; Gutta et al., 2000), and some digital signage systems have recently started to use a fixed video camera for evaluating the content from the viewers. However, those systems mainly gather human attributes which do not necessarily relate to their feelings, such as gender, age and viewing time (Müller et al., 2009). These systems cannot distinguish whether the viewer is watching with negative or positive feeling towards the content, and evaluations are not sufficient to analyze which section of the content was attractive or not.

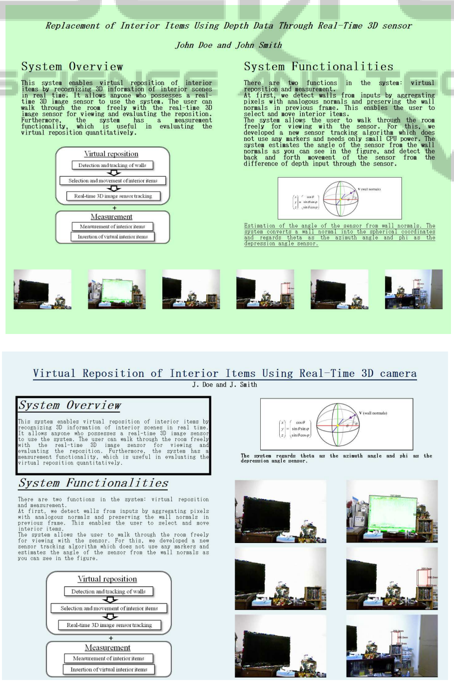

Figure 1: An example of adaptive conference poster. Upper: An initial design. Lower: An evolved design after several generations were altered.

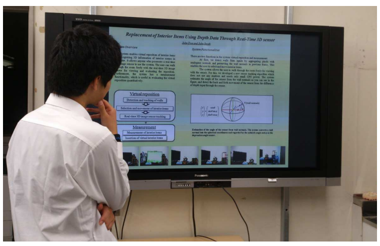

Figure 2: A snapshot of the proposed system in actual use.

From the aspect of aesthetics, the question on the attractiveness of design has been a difficult question because it involves many subjective issues in psychology, statistics and other disciplines. There exist some research challenges that tackled this problem. For example, Singh and Bhattacharya (2010) proposed an algorithm to improve the aesthetics of web interface utilizing a GA for evaluating more than ten geometrical features, such as balance and equilibrium, of page layout. However, the approach does not take user evaluation into primary account, and they concluded that it is difficult to define the cases in which the approach works well.

Moreover, in the case of digital signage, we have to take into account the fact that the local tastes toward the content may be different depending on the place where the digital signage is installed. There are some known digital signage systems which enable interaction with the viewers, where the viewers can be provided the content which they want to see, while the systems have to wait for the explicit interaction from the viewers. Most pedestrians may walk with other purposes and just pass over the digital signage, and therefore it would be inefficient to wait for the interaction from them. In order to ameliorate these problems, we will present a novel approach to make the content more attractive to the viewers by utilizing their "involuntary behaviors" for evaluating the content.

3 APPROACH

Figure 3 shows the processing flow of the proposed method, whose advantages compared to previous works can be summarized as follows.

In the proposed method, the system automatically learns "what content is attractive" and modifies the appearance of the content, so we do not need to analyze the local tastes, search for attractive designs, nor modify the appearance of the content manually. In general, GA is useful for solving problems whose structure is not well understood and for which no exact solution is found. The problem of seeking "what content is attractive" can be regarded as a typical one of such problems, and GA has an ability to overwhelm the manual design.

As for the evaluation, the proposed method utilizing viewers' involuntary behaviors can make it easy to collect many useful evaluation samples. Since digital signage targets the public, a large number of people can serve as viewers, without waiting for their voluntary interaction. In addition, "involuntary behaviors" are deeply related to their feeling, we can draw reli-

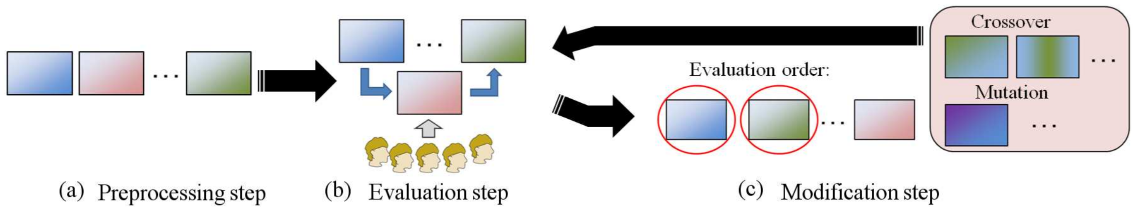

Figure 3: The processing flow of the proposed method. (a) First, the system requires us to define what we want to show to the public and puts up multiple content designs based on the inputs. (b) The system shows each content design in turn, and when each content design is displayed, the system recognizes involuntary behaviors of the viewers in order to evaluate the attractiveness of each content design from them. (c) After all content designs are evaluated, they are sorted in an evaluation order. Then the system breeds new content designs through crossover between highly-ranked content designs and mutation of the highest-ranked content design. After that, the system shows the new content designs once again for the viewers to re-evaluate them. Repeating these steps, the system evolutionally produces more attractive content designs. If an acceptable content design is reached, the evolution terminates.

able evaluation towards the content from them, without any need to check up the validity of the evaluation. "Involuntary behaviors" include gaze-related information such as head pose and pointing fingers, and thus the system can understand which section of the content is attractive.

As shown in Figure 3, the proposed method consists of three steps: (a) *preprocessing step* for making sections from the content of digital signage and creating the first generation of the content design; (b) *evaluation step* which uses the viewers' involuntary behaviors; and (c) *modification step* which produces new generations for making the content more attractive based on the evaluation. After a modification step, the system will gather evaluation to each of the individual content designs utilizing the viewers' involuntary behaviors once again. Repeating these steps, the system evolutionally produces attractive designs of content. If an acceptable content design is reached, the evolution terminates.

Figure 1 illustrates one example of evolved content designs. Each of the figures in the poster was independently adjusted in terms of size, and we can easily identify each section and understand the overall content.

3.1 Preprocessing Step: Extracting Sections

First, we have to make it clear what we want to show to the public. We are required to divide the content into smaller sections. An academic conference poster can actually be divided into "title", "authors", "section titles", "content of subsections" and "reference figures". Hereafter, we will simply refer to the overall design of the content as *content design*, and small content sections defined in this step as *sections*.

Note that there can be found hierarchical relationships among these sections. For example, "title" has "section titles" as its children, and "section titles" are siblings. We have to specify these relationships explicitly in extracting sections from the content.

Each section is required to have two kinds of properties: *content properties* and *graphical properties*. Content properties are what we want to tell to the public in the corresponding section. For example, in the section of "authors", content properties are specified as the names of the authors. Here, we define what values the content properties can take, because there exist various ways to indicate the message. If it is "authors", we make up several patterns of sentences as the defined values for the content property (Figure 4). On the other hand, graphical properties are about how to decorate the section, such as font, font color, frame color and size of the section. These values are given to each by the system automatically.

Next, the system randomly creates a fixed number of initial content designs based on pre-defined sections, giving random property values to graphical properties and randomly selecting the pre-defined values for content properties. Note that the system takes the defined hierarchical relationships between sections into account for unified design. Figure 5 shows one example of the created content design with its underlying hierarchical structure. Such content designs constitute the first generation in the GA.

3.2 Evaluation Step: Recognition of Viewers' Involuntary Behaviors

In this step, the system gathers the viewers' evaluation towards the content designs by recognizing their involuntary behaviors.

3.2.1 Two Behavior Types

In front of digital signage, the viewers behave in various ways. Our system uses Microsoft Kinect™ for recognizing the viewers' behaviors, and in order to make these behaviors easy to understand, we define two types of behaviors: behaviors indicating viewers' attention point (*attention pointer*) and behaviors indicating viewers' feeling (*feeling indicator*).

Attention pointers are behaviors such as changing head angle and pointing gesture. Recognizing these behaviors and considering the position of the viewer and the size of digital signage, the system can estimate which section is focused on. In the current system, eye location is not recognized due to the limited resolution of Kinect. There are many known attempts to utilize eye location for gaze estimation (Valenti et al., 2012). If digital signage is large, we can estimate the attention point only from the head angle, but in order to obtain more accurate results we will rely on the recognition of eye location as hardware technologies develop.

On the other hand, feeling indicators include gestures such as folding arms, nodding, and shrugging shoulders. Such behaviors are deeply related to the viewers' emotions. Finding relationships between these behaviors and feelings has been studied extensively. For example, Gunes and Piccardi (2006) created a bimodal face and body gesture database for analyzing human nonverbal behaviors. The system

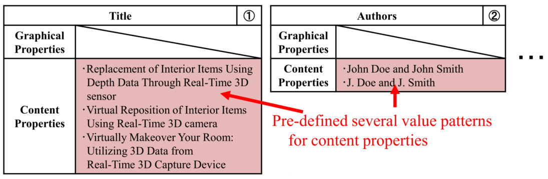

Figure 4: Examples of sections of the content. Each section has content properties and graphical properties, and values are pre-defined for each content property.

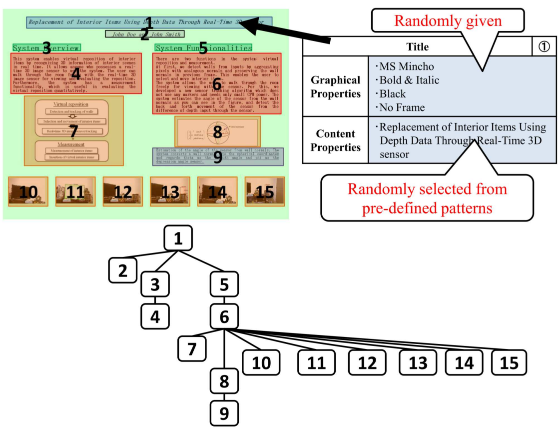

Figure 5: Examples of the properties with the hierarchical structure of the content design created in the preprocessing step.

can learn the meanings of behaviors referring to the database, and by combining this information with the above-mentioned attention pointers, the system can gather the viewers' evaluation towards each section of the content design.

When recognizing these behaviors, several aspects of the diversity deserve to be discussed, as various kinds of people can serve as the viewers for digital signage. One aspect is moving speed of viewers. For example, if the viewer is a child, he behaves quickly, while old people behave slowly. We have to take into account the difference in moving speed and categorize each evaluation. Another aspect is that the meaning of gestures can be different if the viewers' cultural backgrounds are different. For instance, for most cultures, shaking the head side to side implies negative reply, while it implies positive reply for some cultures such as Indian. Thus, taking into account these diversity is important for correct evaluation. In our system, as we can capture whether the viewer is a child or an adult, we change the capturing speed if the viewer is a child. As for the meaning of gestures, keeping in mind that previous research on the detailed meaning of gestures has limited capability because human affection is a subjective matter as mentioned before, we focus only on one culture, Japanese, and then classify the meanings of gestures just according to whether the gesture is positive or negative.

Moreover, we have to consider that the system has to capture the viewers' behaviors without making them aware of it, or their behaviors would be biased. Viewers' behaviors should be "involuntary" and come directly from their actual feelings. Therefore, in addition to the necessity of hiding the capture device, the device should not be mounted and capture the viewers behaviors in a noncontact way. As the device for capturing the viewers' behaviors, our system employs Kinect, which allows us to capture various behaviors such as head pose, facial expression and gestures of multiple viewers involuntarily.

3.2.2 Evaluating the Content Design from Three Aspects

"Evaluation" is considered to form three points of view: the overall evaluation of the content design and the evaluation indicating eye-catching and unconvincing sections in the content design. Each evaluation will be used for making the content more attractive utilizing the GA (see Subsection 3.3 for more details).

The first evaluation is an overall evaluation and used for "selection" in the GA. We define the evaluation as the total of time duration for which the viewers are looking at the content design without negative feeling. We will refer to this evaluation as *overall* *evaluation* of the content design.

The second evaluation is the evaluation which indicates eye-catching sections and used for guiding "mutation" for graphical properties in the GA. We define the evaluation as the total of time duration for which the viewers are looking at the section no matter how they behave. This evaluation indicates how eye-catching the section is, without reference to its content properties. We will refer to this evaluation as *eye-catching evaluation* of the section.

The third evaluation is the evaluation which indicates unconvincing sections and used for guiding "mutation" for content properties in the GA. We define the evaluation as the total of time duration for which the viewers are looking at the section with negative feeling. We will refer to this evaluation as *unconvincing evaluation* of the section.

In this step, we have to take the content of digital signage into account. Viewers' ways of viewing will be substantially different between when the content is an advertisement and when it is a restaurant menu, and we have to consider these differences very carefully. In our system, the content is limited to academic conference posters, and the viewers will take two types of approaches for reading the content: overviewing the whole content and concentrating on reading a particular section. We have to evaluate the gestures separately according to these two approaches, and thus when the viewer does not move his head or he stands close to the digital signage, our system presumes that he is concentrating on reading a particular section and the section should be evaluated more compared to when they are overviewing the whole content.

3.3 Modification Step: Making the Content Designs More Attractive Utilizing GA

After the evaluations, the system sorts them in an evaluation order using *overall evaluation*. Then, the system modifies the content design utilizing the GA. In this step, we have to consider the following two points in addition to the relationships between each section.

The first point is that the modification of the content design should not be done instantly because user experience would be reduced if the system changes its content design when some other viewers are still looking at it simultaneously. Changing the content lazily can be one promising solution for this problem, but it would make the evaluation for individual content designs vague because each of the content designs would not be fixed. Thus, in our system, the content design will switch to another when a fixed time duration runs over and no viewer is looking at it.

Secondly, we should not change the property values of each section in an equally-weighted manner, because the weight of information each section has can be different. For example, the weight of information "title" is much heavier than that of "section title", and changing the font color of title extensively affects the look and feel of the content design more than changing that of section title. We have to consider the weight of information when modifying the content design, and thus our system makes the modification to the section with heavy information such as title section more lazily.

The modification is performed in two ways: *crossover* and *mutation*. Figure 6 shows examples of the breeding.

3.3.1 Crossover

Top two ranked content designs are selected for crossover. In crossover, the system breeds new individuals by combining property values of each section in the highest-ranked content design and the second highest-ranked one. The possibility of values in these combination is based on each *overall evaluation*.

In Figure 6, a new content design whose properties come from the two previous content designs was bred by the crossover.

3.3.2 Mutation

Mutation is applied to the highest-ranked one. In this step, the system breeds some new individuals

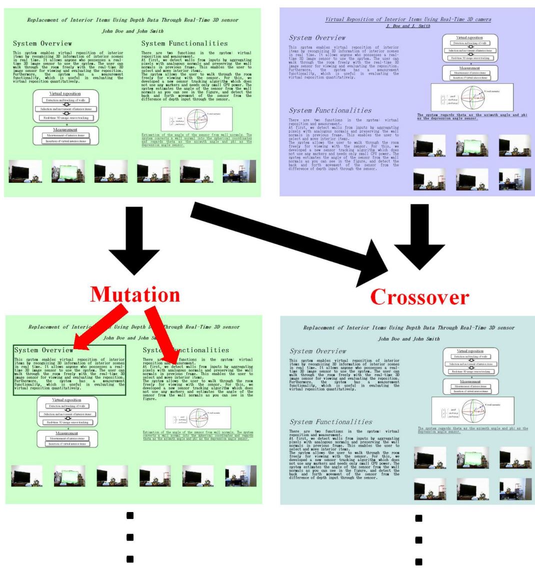

Figure 6: Examples of crossover and mutation of the designs.

by changing the property values of each section in the highest-ranked one. In the system, mutation is not completely executed at random. We use *eye-catching evaluation* and *unconvincing evaluation* for the guide of mutation. The section which got the highest value of eye-catching evaluation is the section which attracted the viewers most. Therefore, to emphasize the section is important for making the content more attractive, and thus the system randomly changes its graphical properties. The system does not need to consider what kind of modification should be done to the graphical properties for attractiveness, because unattractive individuals will be exterminated later. The section which got the highest value of unconvincing evaluation is the section the matter of which viewers thought unconvincing, and thus changing the content properties of the section is necessary for making the content easy to understand, especially in the case the section was the one which got the high value of eye-catching evaluation. In the same way as before, the system does not need to consider what kind of modification should be done to the content properties for making the section more convincing. The system randomly selects the value from the pre-defined values of the content properties. It is because unconvincing sections will be selected for mutation later again. Considering these guides from the evaluations, the system breeds some new individuals by mutation in addition to those by crossover. The system does not apply mutation to the crossovered content designs because we cannot identify the guide for mutation of the crossovered content.

In Figure 6, the section about system overview was emphasized because it gathered much attention, and the sentences of the section about system functionalities were replaced with other pre-defined sentence patterns because they were thought to be unconvincing.

4 CONCLUSIONS

In this paper, we have presented a novel method which does not rely on manual evaluation and automatically makes the content of digital signage more adapted to the local tastes. We take advantage of viewers' involuntary behaviors in front of the digital signage for evaluating the attractiveness of the content, and make the content design more attractive to the viewers utilizing a genetic algorithm, which is useful for solving problems for which no exact solution is found. We empirically proved the feasibility of the method through the development of a pilot digital signage system for displaying academic conference posters.

The current system is a pilot system and can be extended in many directions. One major point is how the system gathers enough evaluations in places with few people, and/or for short periods. We will seek the best timing of changing the content design, and combination of multiple digital signage devices. Besides, we limited the usage of digital signage for only a particular purpose, but we will be able to generalize the method more extensively, considering various behaviors and modifications. For example, we can consider the use of eye-catching evaluation and unconvincing evaluation also in crossover of the GA, and take advantage of hierarchical relationships between each section of the content design more thoroughly.

ACKNOWLEDGEMENTS

The work is supported in part by a Grant-in-Aid for the Leading Graduate School Program for “Science for Development of Super Mature Society” from the Ministry of Education, Culture, Sports, Science and Technology in Japan.

REFERENCES

F. Alt, S. Schneegas, A. Schmidt, J. Muller and Nemanja Memarovic: “How to Evaluate Public Displays,” *Proc. of the 2012 International Symposium on Pervasive Displays*, Article No. 17, 2012.

H. Gunes and M. Piccardi: “A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior,” *Proc. of the 18th International Conference on Pattern Recognition*, pp. 1148–1153, 2006.

S. Gutta, J. R. J. Huang, P. Jonathon and H. Wechsler: “Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces,” *IEEE Trans. on Neural Networks*, Vol. 11, No. 4, pp. 948–960, 2000.

J. Müller, J. Exeler, M. Buzeck and A. Krüger: “ReflectiveSigns: Digital Signs That Adapt to Audience Attention,” *Proc. of the 7th International Conference on Pervasive Computing*, pp. 17–24, 2009.

M. Pantic and Leon J. M. Rothkrantz: “Automatic Analysis of Facial Expressions: The State of the Art,” *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp. 1424–1445, 2000.

N. Singh and S. Bhattacharya: “A GA-based Approach to Improve Web Page Aesthetics,” *Proc. of the First International Conference on Intelligent Interactive Technologies and Multimedia*, pp. 29–32, 2010.

R. Valenti, N. Sebe and T. Gevers: “Combining Head Pose and Eye Location Information for Gaze Estimation,” *IEEE Trans. on Image Processing*, Vol. 21, No. 2, pp. 802–815, 2012.