Design of Interactive Emotional Sound Edutainment System

Myunjin Park and Kyujung Kim

School of Media, Soongsil Univeristy
1-1 Sangdo-1-dong, Dongjak-gu, Seoul, Korea
{cooljin,kyu}@ssu.ac.kr

Abstract. This paper introduces an emotional sound edutainment system for children to learn basic musical composition called as musical education sound interactive game (MESIG) employing a new type of user interface. Developed interactive game interface provides children to enjoy the game, so that they learn how to compose musical notes with touching the tangible objectives instead of using ordinary input devices. This way on experiencing and playing the computer games has been evolved to use the body and hands' movement so as to interact with the game in virtual environment, which brings out interest for the children and their learning capability becomes more effectively improved. This system introduced in this paper requires a single camera and carries out skin color model tracking function to detect hand gesture as input device for playing the game. This computer vision technique based on image processing makes possible to operate an expressive interactive musical education system. To exploit the effectiveness, evaluation and analysis works are accomplished upon the realization of sound edutainment game.

Keywords: Interactive games, edutainment, skin color model, computer vision.

1 Introduction

Computer game plays an important role of education and entertainment to make the children experience more meaningful. Game is a method for shaping children's experience and helping them construct meaning for themselves. The interactive game is a mixture of technology, education and entertainment, so it is evolving within rapidly developing of digital environment. The theoretical of play is as a core activity in a child development and learning. It is also as an essential characteristic of virtual reality environments for interactivity.

Therefore, it is encourage a positive research towards new technologies that offer in reshaping the way in which children interact and present in the game. So it continues in time-honored role in constructing meaning for children [1]. Traditional computer games are mostly played alone at computer or players participate in networking online gaming. The user interface consists of monitor, as well as mouse, keyboard or joystick to control the game. Human gestures are not fully utilized as a sort of interface in the manner of controlling the games [2].

J.A. Jacko (Ed.): Human-Computer Interaction, Part IV, HCII 2009, LNCS 5613, pp. 368–377, 2009.

The developments of interactive and participatory environments that combine the physical and virtual have brought as a natural continuation to the computer game industries. This development area not only restricts to entertainment domain but indirectly creates a learning environment for children. A critical review of examples of immersive virtual reality worlds created for children, with particular attention given to the role and nature of interactivity, is attempted. Interactivity is examined in relation to learning, play, narrative, and to characteristics inherent in virtual reality, such as immersion, presence, and the creation of illusion [3]. Usual virtual reality games try to make the user interface transparent or invisible, where user does not realize the existing of interface to immerse into the games. The conventional method, the player has to use special devices such as gloves or head-mounted displays [4][5].

This motivates us to study a simple manner of children to interact with interesting game contents in interactive digital environment without any bulky device. The interactive digital environment is referring as a system which adapts to the user's action and allows varied degrees of freedom to control over the time or space [6]. The study of the user interaction in virtual environment based on activity theory framework [7] is analyzed and decision to design activity contents to generate larger motivating in game. In this paper, we introduced a tangible input to interact with the virtual environment. Human motion interacts with the virtual environment to gain the children attention and interest to play the games. This interactive games interface able to develop and gain children attention, interest, and learning how to compose simple musical notes throughout their gaming time experience.

The proposed MESIG System is described in section 2 including the explanation of the details of how to track user's body and hand movement and the process for generation and rendering snowflakes and musical notes object by using DirectX. In the last subsection of section 2, the way how to perform the collision detection is described. Section 3 shows the experimental results. The conclusion is summaries in section 4.

2 Proposed MESIG System

The proposed MESIG system provides emotional playing activity game to user, such that appropriate sounds relevant to action of user's body and hand and snow flake movements. The functions of the proposed system can be categorized into three actions. As the first, the image of snow falling is provided to the passengers drawing their attractions if there is no one in front of camera. As the second action, when the specific user stands at the camera, all the snowflakes turns their shape into musical notes having their own colors such that the user feels like to touch. As the last action, by tracking user movement, when the user's hand touches the musical notes, the collided snow flake is disappeared and the corresponding sound classified by its color is emitted. The procedure for processing of this system can be summarized in Figure 1 as shown below.

The overall system determines whether or not there is a person in front of screen. Towards this, the image captured by camera should be changed a binary image using a background subtraction method and recognized user's body and hand using a

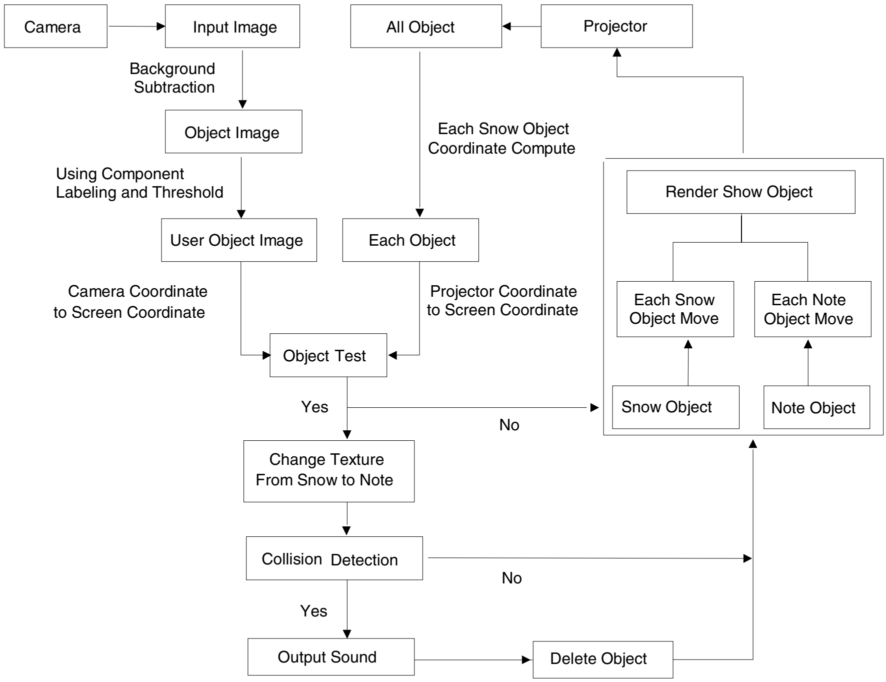


Fig. 1. Overall structure of the proposed system MESIG

component labeling method. As soon as the existence is confirmed, all of the snowflakes subsequently changed into musical notes. Then, utilizing the position information about regions associated with formerly extracted hand and face, the collision detection is performed if the user's hand touches the projected musical notes. Whenever the collision is detected, corresponding musical note becomes disappeared with emitting the sound, at sequel, the new object is appeared. Surely, repeating this process provides the interactive environment to the user controlled by tracking and moving user's hand, so that the snowflake is rendered at the instant of sensing the touch of snowflake with the help of DirectX engine.

2.1 How to Track User's Movement

Since the user described as a kind of interface, when the user touches snow flake, its shape turns out to be musical notes, and relevant sound is generated. Here the importance is how to track the user movement. Towards this, this system uses background subtraction technique in order to segment out objects of interest from captured image in real-time.

Objects are generated using background subtraction and shadow detection based on color difference [8]. From the objects images, we can change from snow flake objects to musical nodes and the musical is disappeared. We used threshold value of RGB color to subtract the background and foreground images. Figure 2 shows the silhouettes pixels extraction based on RGB color. The object of interest generation uses two methods. First, we use the distance in order to separate an object pixel and a background pixel, and the relevant equation for the distance is the following;

$$d(i,j) = \left\| c_r(i,j) - c_b(i,j) \right\| \tag{1}$$

where $\|c\|$ represents the norm of a vector c. Second, we use difference to separate shadow can be expressed by

$$\theta = \cos^{-1}\left[\frac{c_r(i,j) \bullet c_b(i,j)}{\|c_r(i,j)\| \|c_b(i,j)\|} \right] \tag{2}$$

where · is the dot product operator.

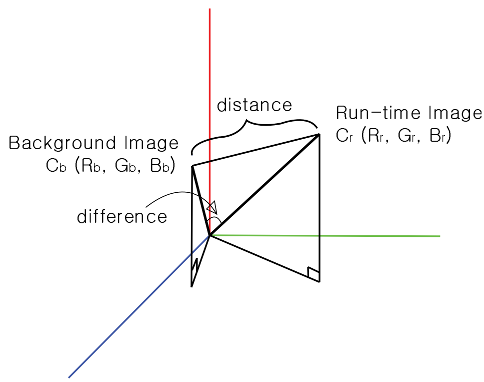


Fig. 2. Overview of objects pixels extraction of *(i,j)*th pixel color

We define a background image whose (i,j)th pixel color is denoted by the vector C_b(i,j) or $C_b(R_b, G_b, B_b)$, and the run-time image whose (i,j)th pixel color is denoted by the vector C_r(i,j) or $C_r(R_r, G_r, B_r)$. The distance is computed by using intensity difference between the run-time and background pixels. If the intensity difference is very large, then the pixel should be a silhouette pixel. If the intensity difference is very small, then the pixel must be a background pixel. Shadow pixel is determined by performing pixel determination to exclude silhouette and background pixel as described above. The value difference is the angle between the vectors C_r and C_b in the RGB color domain and hence is a measure of the color difference between the run-time and background pixels in Figure 3. After background subtraction, the extracted object is shown in Figure 3. The noise could be appeared due to musical notes and snowflakes when real-time captured image is extracted as shown in Figure 3(b)-(c) based on background subtraction. In order to discriminate the noise and the person, each component is classified into the pure noise and the person by comparing those with the threshold value. Here, it is easy to determine whether the component is

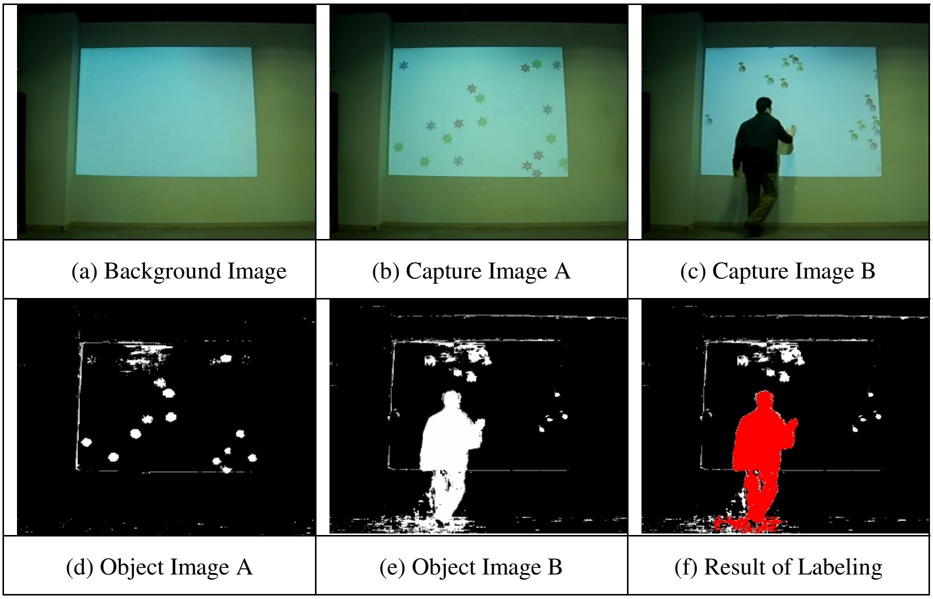


Fig. 3. MESIG of user tracking

the noise or the person since the component corresponding to the person has more large number of pixels. In the proposed system, the threshold is fixed to 1200, so that the component is recognized into the person if the number of pixels is more than 1200, otherwise the noise. Figure 3(f) is the result after the discrimination and the suppression of noise from the subtraction image shown in Figure 3(e), and the shaded component is recognized as the person.

2.2 How to Perform Generation and Rendering Snow Flake Object Using DirectX

This section utilizes DirectX as tool for generating realistic snowflakes or musical notes. It is well-known that DirectX provides comfort API to control the functions of high performed hardware. Thus, more powerful multimedia applications can be easily developed. In order to project realistic snowflakes or musical notes to the user, the eight steps are processed as shown in Figure 4 by using DirectX.

First, the user sets up the total number of snowflakes. If the number of snowflakes is not enough as determined, residual snowflakes are generated by random selection of the number of snowflakes in random. After performing this step, as the second step, all the snowflakes has its own projection coordinate, and its magnitude and moving path are determined one by one. Each snow flake object follows its own moving path. As the third step, the presence of the user is testified. At fourth step, all the snowflakes turn their shapes into musical notes. Then, the collision detection is processed whether there is any touch between hand or face and musical notes in fifth step. If there is any collision, the corresponding musical note emits a sound and a color in sixth step. Here, each musical note can have 7 colors or sounds. The collided

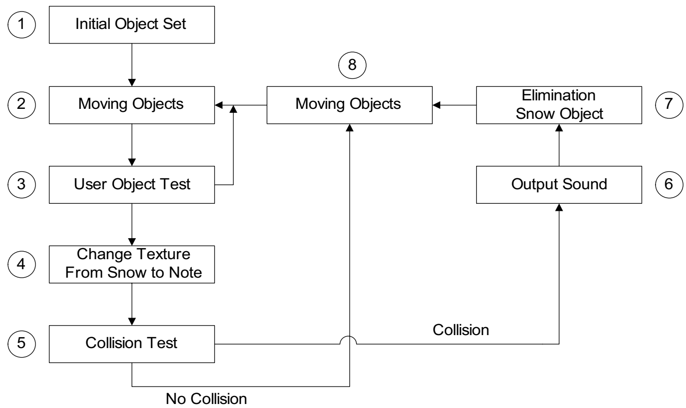


Fig. 4. Processing steps for the proposed MESIG system

musical note becomes disappeared in seventh step, whereas the object without collision keeps moving its position in eighth step. After the sound note is eliminated, the number of snowflakes will be decreased by the number of the disappeared. These 1 to 8 steps are repeated until all the snowflakes are disappeared. Figure 4 shows these processing steps for generation, movement, change and elimination of snowflakes.

2.3 How to Perform Collision Detection

It is important to detect the event that snowflakes collide with user movement. Since the user performs interaction by using body or hand movement, this interface can provides very interesting situation. In order to perform the collision detection, the functions for tracking the user's hand or face explained in Sec. 3 and the localization of projected snow flake explained in Sec. 4 should be performed. However each snow flake has its own coordinate, it hard to perform the collision detection [8] by only using the location of projected snowflakes. In other words, the coordinate oriented from the camera tracking the user's hand or face is totally different from that from projector for snowflakes, it is necessary to use the common coordinate. This common coordinate can be updated by transforming the coordinate of projector into that of screen. After all, the collision detection is performed by this coordinate transformation. Since the coordinate of screen can be calculated by the projected region, it can easily test whether the collision happens or not.

To make these coordinate changes, the keystone correction technique can be generally utilized for providing the corrected image wherever the image is projected from the projector with considering relative positions between the camera and the projector. Here, in order to detect the collision, the following equations (3) and (4) are used to transform the coordinates of the camera and the projector into the common screen coordinate. First of all, the relationship between the coordinate of projector and that of camera is designated as in LHS of equation (3) with whose corresponding transformation matrix T. And secondary, the relationship between the coordinate of screen and that of camera is defined as in RHS of equation (4) by using C. Using these relations, the relationship between the coordinate of screen and that of projector can be expressed in equation (4), whose relevant transform matrix defined as P

$$T \times \begin{pmatrix} x_{proj} \\ y_{proj} \\ 1 \end{pmatrix} \equiv \begin{pmatrix} x_{cam} \\ y_{cam} \\ 1 \end{pmatrix}, \quad C \times \begin{pmatrix} x_{world} \\ y_{world} \\ 1 \end{pmatrix} \equiv \begin{pmatrix} x_{cam} \\ y_{cam} \\ 1 \end{pmatrix} \tag{3}$$

$$C^{-1} \times T \times \begin{pmatrix} x_{proj} \\ y_{proj} \\ 1 \end{pmatrix} \equiv \begin{pmatrix} x_{world} \\ y_{world} \\ 1 \end{pmatrix}, \quad P = C^{-1} \times T \tag{4}$$

Utilizing equations (3) and (4), the coordinate change from camera to screen can be accomplished, and Figure 4 visually depicts these coordinate transformations step-by-step. From the results of these transformations, the collision detection is performed by matching the coordinate of snow flake with that of either user’s hand or face.

3 Experimental Results and Analysis

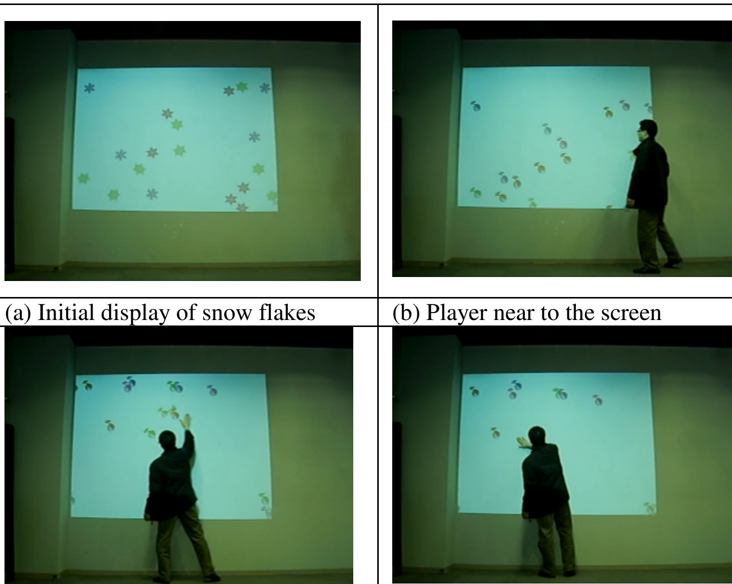
In this section, we present result of the proposed game MESIG. The proposed game required a camera (LG – color CCD camera), a projector, screen and a personal computer (Intel® Xeon™ CPU 3.20GHz, 2.56GB RAM, Windows XP OS) to execute the MESIG. The display screen will be change when the real-time camera of the system detect human or hand skin color. Then the musical notes display will replace the snowflakes display. Figure 5 (a) shows the overall system setup of MESIG which is projected to a wide screen; (b) shows the initial display of snowflakes when no player detected in real-time camera. The display screen environment will change from snowflakes to musical notes when a player face or hand skin color is detected. The musical notes have 7 basic elements of sound notes which are represented in rainbow colors. When player hand touches on a particular note, the sound notes will be generated. Figure 6 shows the result of MESIG with different players captured from the normal digital camera. The system shows the interactive game environment by detect the player skin color. When the player’s hand touch on the musical notes, a sound will be generated based on the notes representation. This MESIG will motivate the children interest and creativity to compose musical notes. The sound notes only response to one time hand touch of the musical notes.[1]

This analysis is accomplished for 30 children as the examination group whose age ranges from 4 to 7. The pre-exam is conducted before the children play, then after the post-exam is executed to analyze the degree of effectiveness. In order to determine how the educational interest and achievement is improved, the basic questionnaires are assigned to the children, which are including the preference and the skill about

[1] www.cooljin.net/conf08/iadis_elearing

(a) Overall system | (b) Initial display of snow flakes

Fig. 5. Proposed MESIG system environment



(a) Initial display of snow flakes | (b) Player near to the screen

(c) Player1 touch the notes | (d) Player2 touch the notes

Fig. 6. Experimental result shows the player plays the MESIG using hand to touch the musical notes to generate a sound notes

computer games. For fair evaluation, the corresponding questionnaires are selected from the pool made by the people who have the major on child education. Experimental results are subsequently analyzed by conducting SPSS program. Among those, firstly, the sound educational achievement is exploited by evaluating T-value a priori. At second, after the end of executing game, to concrete the inference a priori made, the posteriori T-value evaluation process is conducted at sequel. Significance level of all the experiments is fixed, whose corresponding value is less than 0.05. Table 1 shows the results of evaluation before and after play of the sound edutainment game.

Table 1. Level of interest for sound edutainment game

Subject	Level	Examination Group (30 Children)	
		Pre-evaluation	Post-evaluation
		N(%)	N(%)
Sound Interest	High	8(27%)	13(43%)
	Middle	11(36%)	12(40%)
	Low	6(20%)	3(10%)
	No Response	5(17%)	2(7%)

Table 2. Achievement of sound perception education

Subject	Evaluation	Num. of Student (N)	Average	Standard Deviation	T-value	P-value
Sound perception	pre-examination	30	11.47	3.42	5.567	0.01
	post-examination	30	13.84	2.65		

As shown in Table 1, before the game play, the level of interest is not so high, but after the game play the level of interest is positively improved. In Table 1, among examination group, it can conclude that the number of children having high level of interest is increased from 27% up to 43%. Besides the increment of interest, the capability of sound perception is also enhanced. The level of achievement corresponding to the children in examination group is improved by 2.37 points. According to $t=5.567$ and $p=0.01$. From the statistical point of view, the results in Table 2 can be said to be meaningful because $p < 0.05$. And the results show that the sound perception education via the interactive game is quite effective to the children.

4 Conclusion

This paper introduced an interactive game for children to learn basic musical composition. The proposed system is an interactive game named as musical education sound interactive game (MESIG), which can be treated as a new type of user interface. Developed system is designated an interactive game interface for children to enjoy and play the games while learning how to compose musical notes using tangible input instead of ordinary input devices. Thus, the proposed MESIG system provides emotional playing activity game to user, such that appropriate sounds corresponding to action of user's hand and snow flake movements. The functions of the proposed system can be categorized into three actions. To verify the effectiveness of the proposed game for sound perception edutainment, the fair statistical analysis is conducted. All the results show the improvement of sound perception capability as well as the increment of interest. This indicates that the interactive game controlled by body-motion can strongly motivate the reason for learning, so that this new way of education via edutainment in a form of interactive game increases the level of learning interest and improves the capability of sound perception more effectively.

Acknowledgments. This work was supported by Soongsil Research Fund.

References

1. Madej, K.: Towards digital narrative for children: from education to entertainment, a historical perspective. Comp. in Entertainment (CIE) 1(1) (2003)
2. Stromberg, H., et al.: A group game played in interactive virtual space: design and evaluation. Symposium on Designing Interactive Systems (2002)
3. Roussou, M.: Learning by doing and learning through play: an exploration of interactivity in virtual environments for children. Computer in Entertainment (CIE) 2(1) (2004)
4. Bobick, A., et al.: Perceptual user interfaces: The KidsRoom. Communication, ACM 43(3), 60–61 (2000)
5. Pausch, R., et al.: First Steps Toward Storytelling in Virtual Reality. In: Computer Graphics Proceedings, ACM SIGGRAPH. Annual Conference Series, pp. 193–203 (1996)
6. Talin: Real interactivity in interactive entertainment. In: Dodsworth Jr., C. (ed.) Digital Illusion: Entertaining the Future with High Technology. Addison-Wesley, Reading (1998)
7. Nardi, B.A.: Context and Consciousness. In: Activity Theory and Human-Computer Interaction. MIT Press, Cambridge (1996)
8. Kong, G., et al.: A real time system for robust 3d voxel reconstruction of human motion. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition., vol. 2, pp. 714–720 (2000)
9. Sukthankar, R., et al.: Smarter Presentations. Exploiting Homogra-phy in Camera-Projector Systems. In: Proceedings of International Conference on Computer Vision, vol. 1, pp. 247–253 (2001)