Proposal of Interactive Projection Mapping using Human Detection by Machine Learning

Takahiro Shinoda, Makoto Sakamoto*, Takahiro Ishizu, Kenji Sakamoto and Amane Takei
Faculty of Engineering, University of Miyazaki,
Miyazaki-City, Miyazaki, Japan

Takao Ito
Graduate School of Engineering, Hiroshima University
Higashi-Hiroshima, Hiroshima, Japan
E-mail: hm14021@student.miyazaki-u.ac.jp, fruits2000jp@yahoo.co.jp
*Corresponding Author
E-mail: fruits2000jp@yahoo.co.jp

Abstract

In recent years, "Entertainment Computing" (EC) has attracted attention and has become one of the major industries in Japan. "Projection Mapping" is well known in this EC. Projection mapping is a video technology that creates a new space by synthesizing space and video using a projector. Among them, many people are fascinated by works that create a fantastic world by combining dancer performance and projection mapping. However, these works require the performer to accurately align with the coordinates of the image objects in the projection mapping, which is not easy for everyone. In this study, we aim to entertain not only the people who see the projection mapping but also performers. Therefore, we have prototyped an interactive projection mapping that changes according to user movement. This time, we focused on sports and projected the ball to the user to experience baseball pitching and soccer lifting. Furthermore, we conducted a questionnaire survey to evaluate the sense of use of this system, and the results showed that many people can enjoy by this projection mapping.

Keywords: Entertainment Computing, EC, Projection Mapping, Kinect.

1. Introduction

In recent years, exciting projection mapping has received more and more attention. Many people have seen projection mapping to buildings such as theme parks.
For example, there is a closing ceremony for the "Rio Olympics" held in August 2016.
Currently, projections are also being made on familiar items such as clothes, faces, and notebooks. For example, during a performance at the world's largest advertising festival in Cannes, the artist's white costume became a screen, and colorful graphics were projected one after another.

Projection mapping will delight us in many ways. However, in the conventional projection mapping, the viewer mainly enjoys watching and enjoying the projected image, so that the viewer's feeling of immersion in the content is considered insufficient [1].

In this study, we propose a participatory projection mapping that changes according to the movement of participants by projecting to participants.

Takahiro Shinoda, Makoto Sakamoto, Takahiro Ishizu, Kenji Sakamoto, Amane Takei, Takao Ito

2. Study method

2.1. *Equipment Used*

2.1.1. *Kinect for Windows*

Kinect for Windows is available from Microsoft. It's a peripheral device that enables operations by body movement, gestures, sound, etc. without using a controller.

Kinect has an infrared sensor, an RGB camera that acquires 8-bit 3-channel (RGB) image data, a depth image sensor that acquires distance (depth) image data from Kinect, and a sound microphone that estimates the position of a sound source.

Also, Kinect's feature is "posture recognition technology", recognizes the whole human body and performs operations by its movement. Thereby, it is possible to estimate where the parts of the body are by using the depth image.
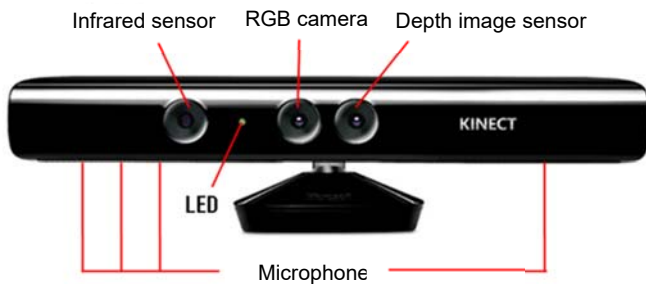

Fig. 1. Kinect for Windows v1 [2]

2.1.2. *Projector*

The projector is used for projection.

2.2. *Development Environment*

- OS: Windows10,
- Development Environment: Visual Studio 2017,
- Development Language: C++,
- OpenNI2,
    OpenNI is an interface that allows you to configure Kinect, acquire images, measure distance, and use a microphone.
    It is also used in this study because it can be linked with OpenCV, an image processing library.
- NiTE,
    NiTE is used to get data (eg gesture recognition, motion tracking etc) from Kinect [2].

- OpenCV,
    Create an object detector using the OpenCV library.
- OpenGL.

3. Outline of the system

It was programmed to display seven screens: Ball, Color, Depth, User, Skeleton, Combination (for projection), and Combination PC (for PC).
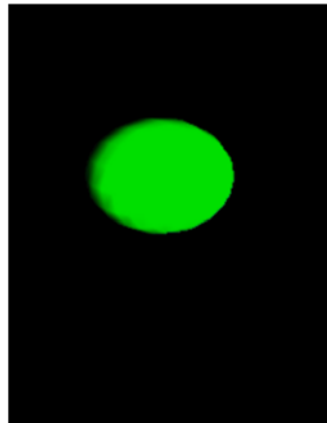

Fig. 2. "Ball" screen display

Fig. 3. "Color" screen display

Fig. 4. "Depth" screen display

Fig. 5. "User" screen display
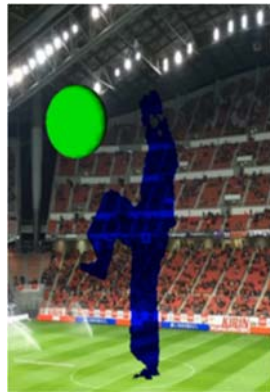
Fig. 6. "Skeleton" screen display Fig. 7. "Combination" ″ Combination PC" screen display

3.1. *User Detection*

3.1.1. *Boosting*

Boosting is a learning algorithm that sequentially generates weak classifiers and combines them to create a strong classifier [3]. In this study, we decided to use a method called AdaBoost among various Boosting methods. AdaBoost is a technique for creating a classifier with high accuracy by learning by adaptively weighting the recognition rate of the classifier during the learning process [4].
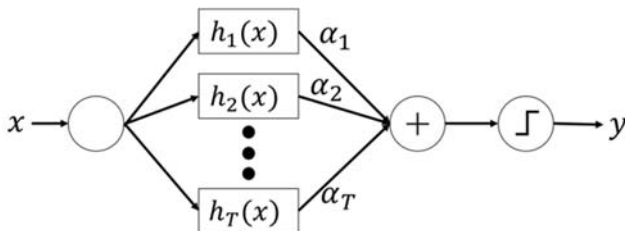

Fig. 8. AdaBoost

3.2. *Extract image features*

The classifier learns by extracting the feature amount of the image at the time of creation. There are the following three feature extraction methods.
- Haar-Like,
- Local Binary Pattern (LBP),
- Histogram of Oriented Gradients (HOG).
In this study, we conducted experiments using LBP.

3.2.1. *LBP*

LBP is one of feature quantities that can be used for image recognition and classification [5]. It is usually calculated in a 3x3 pixel area and extracts local features. It is particularly resistant to lighting changes and has the advantage of being able to calculate at high speed. In the calculation method, first, the luminance value of the center is compared with the luminance values of pixels in the vicinity of the surrounding eight. Of the 8 neighborhoods, 1 is set when the luminance value is equal to or greater than the central luminance value, and 0 is set otherwise. This is multiplied by a mask, the sum is obtained, and this value is replaced with the luminance value of the central pixel. A mask is a weight assigned 2^n clockwise from the top left. This operation is performed for all pixels, and the resulting image is called an LBP image. The object is recognized using the LBP feature value thus obtained.

3.3. *Soccer mode*

The first screen is in the black initial state. When the user raises his knee, a soccer ball and a soccer stadium are projected. In addition, a sound is now played when kicking the ball. If the ball coordinate is lower than the y coordinate threshold, the ball disappears and is initialized. In addition, physics was used for the lifting operation.

Fig. 9. Soccer mode execution result

3.4. *Baseball mode*

The first screen is in the black initial state. When the user raises his hand, a baseball ball and a baseball stadium are projected. In addition, a sound is now played when throwing the ball. In order to make the ball appear to be thrown in the back, the ball gradually gets smaller as the flight distance increases. When the ball position is greater

than the threshold value of the x coordinate (the left edge of the screen as seen by the user), the ball disappears and is initialized.

Fig. 10.   Baseball mode execution result

4.   Experimental result

The experiment using the threshold of skeleton coordinates and the method of this study were tested and compared [6]. In this study, we performed grayscale on the "Color" screen display obtained from Kinect and detected the user using object detector. Figure 11 shows human body recognition in soccer mode.

Fig. 11.   User detection using human detector in soccer mode

In previous study, when the subject disappeared from the Kinect field of view and entered the field of view again, there was a problem that the skeletal information was not retraced and mapping was not successful. In this study, we solved the problem by performing human body recognition using OpenCV without using tracking of skeleton coordinates.

5.   Conclusion

In this study, we used an object detector created with Opencv to create an interactive projection mapping whose mapping changes according to human movement.

In previous study, when the subject disappeared from the Kinect field of view and entered the field of view again, there was a problem that the skeletal information was not retraced and mapping was not successful. In this study, we were able to solve these problems by using a method that does not use arbitrary thresholds. For example, in previous artist concerts, artists have been following the movement of projection mapping. However, with this method, it is possible to perform projection mapping that matches the movement of the artist. This will reduce the difference between each other's movements and make it possible to produce more realistic performances. In the future, we would like to realize interactive projection mapping using more accurate human body classifiers by increasing the number of sample images.

References
1. K.Ogasa, H.Katayose "TPPM (Take Part in Projection Mapping): Multiplayer participation type projection mapping application using tablet terminal" 2014
2. N.Koeda, E.Ueda, Y.Nakamura ″Introduction to image processing with OpenCV″ 2018.
3. Nobuyuki Nishiuchi, Takuma Suzuki, Kimihiro Yamanaka "Development of Work Posture Evaluation System Using Boosting" J Jpn Ind Manage Assoc 62,51-58,2011
4. Yoav Freund, Robert E Schapire "A Decision-Theoretic Generalization of On-Line Learningand an Application to Boosting" journal of computer and system sciences 55, 119139 (1997)
5. Ojala T, Pietikäinen M & Harwood D, "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions.", Proc. 12th International Conference on Pattern Recognition (ICPR 1994), Jerusalem, Israel. Vol I, 582-585. (1994)
6. Makoto Sakamoto, Takahiro Shinoda, Takahiro Ishizu, Masahiro Hori, Hazuki Watanabe, Amane Takei, Takao Ito "A Proposal of Interactive Projection Mapping Using Kinect" IEEE Xplore: 29 November 2018, DOI: 10.1109/ICT-ROBOT.2018.8549899