

Bullet-Time Book: Augmentation of Visual Information in Figures by Bullet-Time Video Display

Qiu Xin Yi

University of Tsukuba
Tsukuba, Japan
biongchou@gmail.com

Hidehiko Shishido

University of Tsukuba
Tsukuba, Japan
shishido@ccs.tsukuba.ac.jp

Yoshinari Kameda

University of Tsukuba
Tsukuba, Japan
kameda@it.tsukuba.ac.jp

Itaru Kitahara

University of Tsukuba
Tsukuba, Japan
kitahara@ccs.tsukuba.ac.jp

ABSTRACT

This paper proposes a method to extend the visual information of books using multi-viewpoint video presentation technology. We capture a book with a camera, from which included figures or photographs in the image are recognized using image features. Based on the recognition result, a bullet-time image registered in the database is searched and displayed. The proposed method makes it possible to observe bullet-time video using a figure in the book as the search key. Bullet-time video enhances a book's visual information because it presents 3D information that is difficult to express in a 2D image, and the viewer can freely determine the observation viewpoint.

Author Keywords

Multiple viewpoint images; Bullet-time video; Visual augmentation

ACM Classification Keywords

Mixed/augmented reality; Graphical user interfaces; 3D imaging; Object identification

INTRODUCTION

With the development of video media technology, book media expression and browsing methods have diversified. Conventional magazines and books consist mainly of text and images; however, electronic books, such as audiobooks and eBooks, can achieve advanced expressions and browsing methods using multimedia information that would be difficult for paper media. In addition, methods of presenting visual information, such as free-viewpoint video and bullet-time video, are being actively researched and developed. Bullet-time video is a method of presenting 3D information based on motion parallax by placing a plurality of cameras around a subject and observing while switching the multi-view video. Introducing bullet-time video for expressing and

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

displaying the pictures and figures in books produces a novel visual experience by expressing 3D information that is difficult to express with a 2D image.

QR code is a well-known method for linking articles of books with multimedia information. By scanning the QR code, it is easy to acquire a link to the multimedia information embedded in the code; however, printing the QR marker impairs the design of the book. Using a figure or picture printed in the book as link information to the multimedia information solves this problem. For example, in MagicBook [1], corresponding link information is detected using captured images of AR (augmented reality) markers on books, thus enabling AR presentation.

In this research, we use image features to recognize figures and pictures contained in images taken of books, and search and reproduce the bullet-time video registered in the database using the results. This method makes it possible to observe bullet-time video with the book's figure as the search key without keyboard input, thus enhancing the book's visual information.

RELATED WORK

There has been considerable progress in information retrieval using image recognition processing and its application system. Maruyama et al. [2] performed object recognition processing using the image feature amount (SURF) and a color histogram for images taken of food, and provided recipe recommendation applications based on the results. Mensink and van Gemert [3] identified the author, type, material, and age of artwork with the recognition process using the SIFT (scale invariant feature transform) feature. Xiong and Ding [10] developed a method to detect generic objects using a single query image without a training process by densely computed SIFT descriptors to extract "gradient distribution" features of the image.

Studies on generation and presentation of multi-view and bullet-time videos are advancing. Akechi et al. [4] proposed a method for generating bullet-time video that resets a point of interest when switching the video's viewpoint. Kitamura [5] improved the visibility of bullet-time video by developing a browsing interface using multi-touch input. Shishido et al. [11] proposed a method for generating pseudo-dolly-in video that can reproduce motion parallax based on 3D estimation from multi-view videos.

Other research aims to expand the visual information of books. Dünser et al. proposed a method of visualizing physical phenomena such as magnetic fields using AR presentation. AR text has been confirmed to increase student concentration as compared to the use of physics textbooks [6]. Vinumol et al. [7] succeeded in enhancing the concentration of students with learning disabilities by presenting images, moving images, and sounds via AR. Grasset et al. [8] are developing children's AR picture books. In this paper, we propose a bullet-time book (BT-Book) that expresses 3D information that is difficult to express with 2D images using bullet-time video presentation to extend the visual information of books and improve browsability.

VISUAL AUGMENTATION OF BOOK BY BULLET-TIME

As shown in Fig. 1, our proposed BT-Book consists of a computer (tablet terminal) equipped with a touch panel interface and a book. We recognize figures and photographs included in the image of the book using the image feature, search the bullet-time video registered in the database using the recognition result, and finally reproduce it on the tablet terminal.

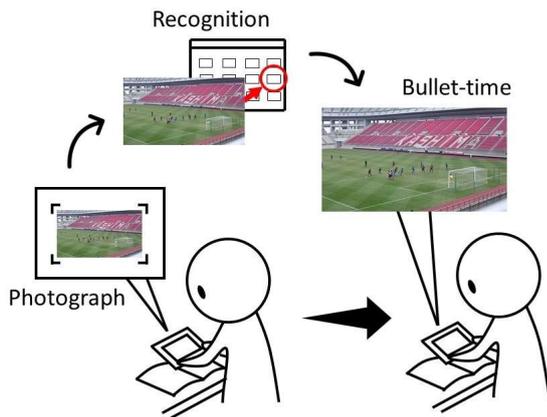


Figure 1. Augmentation of visual information in figures with bullet-time video.

BT-Book consists of four processes of "shooting," "preprocessing," "recognition," and "browsing." First, the user captures a book image (Step 1). In preprocessing, rectangle region detection is performed on the captured image and a target figure region of recognition processing is specified. In recognition processing, image features of the target region are detected using AKAZE (Accelerated-KAZE) [9], and the corresponding bullet-time video is recognized by matching processing with the image features of registered images (Step 2). When the recognition processing is successful, the bullet-time video is called up. If recognition fails, we return to the shooting state (Step 3). The bullet-time video presentation is executed by Kitamura et al.'s multi-touch interface browsing method (Step 4) [5].

BOOK IMAGE SHOOTING AND RECOGNITION

Shooting and Preprocessing

Shooting: In this research, we used Microsoft Surface Pro 4 as the tablet terminal. The Surface Pro 4 has an 8-megapixel rear-camera and 5-megapixel front-camera. For our study, we shoot a 1920 pixels x 1080 pixels image using the rear-camera and save the image after resizing to 0.25 times to reduce the computation cost of recognition processing. An example of a captured image is shown in Fig. 2 (a).

Preprocessing: Preprocessing detects a rectangular region from the saved image. First, as shown in Fig. 2 (b), a target frame region (shown as a broken line rectangle in Fig. 2 (a)) set on the interface screen is cut out (first cutout processing). Next, rectangle detection processing is applied to Fig. 2 (b), and the detected region shown in Fig. 2 (c) is cut out (second cutout). At this time, binarization processing is applied to Fig. 2 (b) by the local threshold method, thereby improving extraction precision. The success or failure of the second cutout is determined by the outline shape and the size of the cutout region. When the second cutout has failed, Fig. 2 (b) proceeds to recognition processing as a rectangular region.



Figure 2. Processing of cutting the photo region.

Recognition

Feature point detection: SIFT, SURF, KAZE, AKAZE, etc. are image features. In this research, we perform feature point detection using AKAZE. AKAZE is an improved algorithm based on KAZE that shortens the calculation time while maintaining detection accuracy. As with SIFT and SURF, it is robust to changes in scale, lighting conditions, and image rotation.

Matching with registered databases: A matching process is performed by a descriptor matcher using a feature descriptor calculated for each image feature point of AKAZE. We use the Euclidean distance in the feature vector space. By setting an upper limit on the distance considered as matching, outliers are eliminated. The matching process adds the

number of feature points determined to match between the captured image and the image registered in the database, and outputs the database with the largest number as the recognition result.

BULLET-TIME VIDEO GENERATION

To capture multiple images for generating the bullet-time video, a plurality of cameras is arranged around the subject to perform synchronous shooting. In this method, 4K-resolution (3840 pixels x 2160 pixels) images are shot at 30 frames per second. Camera parameters necessary for reconfiguring the viewer’s focus point are estimated using Structure from Motion (SfM). We employed an open library VisualSfM in this research.

When a bullet-time video is shown on a tablet device’s monitor, the user can select a focus point by touching any point on the screen. The 3D position is then calculated by applying stereo vision to the other multi-view images. The user can change the observing viewpoint by switching the displayed multi-view image. Figure 3 shows some examples of generated bullet-time video. Our method provides users with high visibility of the target scene.



Figure 3. Bullet-time video.

BT-BOOK INTERFACE

Interface design

The BT-Book interface, which includes a browse-start button (Button 1), a browse-end button (Button 2), a message box, and a presentation screen (entire window), is shown in Fig. 4. A quadrilateral frame to be used in the clipping process of the figure region is superimposed on the presentation screen. When shooting a book, the user is instructed to fit the target (figure or photograph of the book) within the frame.

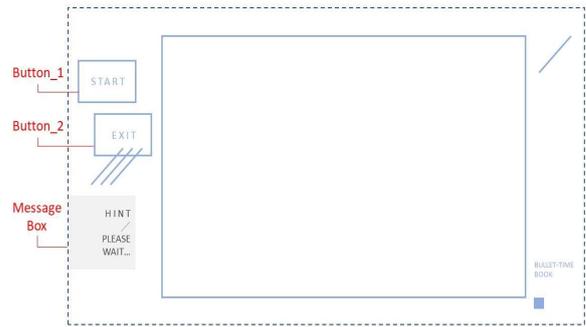


Figure 4. BT-Book interface.

Flow of BT-Book browsing

When the browsing system is activated, a shooting screen as shown in Fig. 5 is displayed on the tablet screen. If the browse-start button is pushed, recognition processing begins after shooting and preprocessing. Thereafter, during the recognition process, a waiting screen is displayed. When the recognition work is successful, the message window closes and the corresponding bullet-time video is retrieved from the database. If recognition fails, the system returns to the shooting screen and urges recapture.

Figure 6 shows an example of scenes using the BT-Book. The user can three dimensionally observe the photo in the document by using bullet-time video.

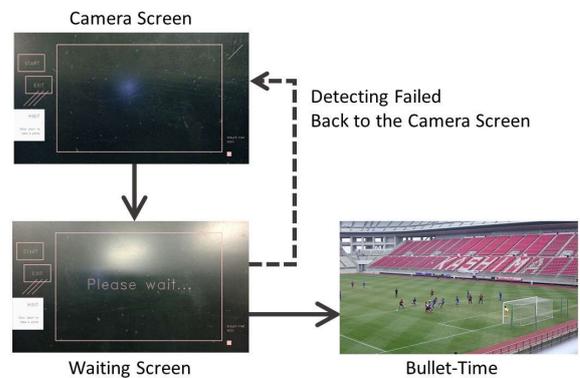


Figure 5. Process of BT-Book interface.



Figure 6. BT-Book in use.

CONCLUSION

In this research, we proposed a method to extend the visual information of books using multi-viewpoint video presentation technology. By implementing a bullet-time video browsing interface using a touch panel, we created a browsing method that enables intuitive viewpoint switching operation with 3D information representation that is difficult to express with 2D images.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 17H01772 and JST CREST Grant Number JPMJCR14E2, Japan.

REFERENCES

1. Mark Billinghurst, Hirokazu Kato, Ivan Poupyrev, "The MagicBook - Moving Seamlessly Between Reality and Virtuality," *IEEE Computer Graphics and Applications*, 21:3, pp.6-8, 2001.
2. Takuma Maruyama, Yoshiyuki Kawano, Keiji Yanai, "Real-time Mobile Recipe Recommendation System Using Food Ingredient Recognition," *Proc. of the 2nd ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices*, pp.27-34, 2012.
3. Thomas Mensink, Jan van Gemert, "The Rijksmuseum Challenge: Museum-centered Visual Recognition," *Proc. of the International Conference on Multimedia Retrieval*, p.451, 2014.
4. Nao Akechi, Itaru Kitahara, Ryuuki Sakamoto, Yuichi Ohta, "Multi-resolution Bullet-time Effect," *SIGGRAPH Asia Posters*, Article No.30, 2014.
5. Nobuyuki Kitamura, Yoshinari Kameda, Itaru Kitahara, "A Method to Switch Multiple-view Videos Using Multi-touch Interface," *Asia-Pacific Workshop on Mixed-Reality (APMR2016)*, pp.22-24, 2016.
6. Andreas Dünser, Lawrence Walker, Heather Horner, Daniel Bentall, "Creating Interactive Physics Education Books with Augmented Reality," *OzCHI '12 Proc. of the 24th Australian Computer-Human Interaction Conference*, pp.107-114, 2012.
7. K.P. Vinumol, Ashish Chowdhury, Radhika Kambam, V. Muralidharan, "Augmented Reality Based Interactive Text Book," *Virtual and Augmented Reality (SVR)*, pp.232-235, 2013.
8. Raphaël Grasset, Andreas Dünser, Mark Billinghurst, "Edutainment with a Mixed Reality Book: A Visually Augmented Illustrative Children's Book," *ACE '08 Proc. of the 2008 International Conference on Advances in Computer Entertainment Technology*, pp.292-295, 2008.
9. Pablo Fernández Alcantarilla, Jesús Nuevo, Adrien Bartoli, "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces," *British Machine Vision Conference (BMVC)*, 2013.
10. Bin Xiong, Xiaoqing Ding, "A Generic Object Detection Using a Single Query Image Without Training," *Tsinghua Science and Technology*, 17:2, pp.194-201, 2012.
11. Hidehiko Shishido, Kazuki Yamanaka, Yoshinari Kameda, Itaru Kitahara, "Pseudo-dolly-in Video Generation Combining 3D Modeling and Image Reconstruction," *IEEE ISMAR 2017 Workshop on Highly Diverse Cameras and Displays for Mixed and Augmented Reality (HDCD4MAR2017)*, pp.327-333, 2017.