

# Interactive e-Book Linking Text and Multi-View Video

Xinyi Qiu  
University of Tsukuba  
Tsukuba, Japan

[qiu.xinyi@image.iit.tsukuba.ac.jp](mailto:qiu.xinyi@image.iit.tsukuba.ac.jp)

Hidehiko Shishido  
University of Tsukuba  
Tsukuba, Japan

[shishido@ccs.tsukuba.ac.jp](mailto:shishido@ccs.tsukuba.ac.jp)

Ryuuki Sakamoto  
Denq Vision Inc.  
Tokyo, Japan

[rkskmt@denqvision.com](mailto:rkskmt@denqvision.com)

Itaru Kitahara  
University of Tsukuba  
Tsukuba, Japan

[kitahara@ccs.tsukuba.ac.jp](mailto:kitahara@ccs.tsukuba.ac.jp)

**Abstract**—This paper proposes an interactive e-book which multi-viewpoint images and text are linked, to let user easily understand the contents such as catalogue or textbook which are composed of images and explanation. By displaying bullet-time video, it makes possible to enhance the visual information of book because it can present 3D information difficult to express in the 2D image, and the user will be able to freely determine the observation viewpoint. Two functions are designed to link the specific keywords of text and viewpoint of bullet-time video. These interactive functions can lead users to read the contents smoothly and quickly.

**Keywords**—Bullet-time, Multi-view images, Multi-media book

## I. INTRODUCTION

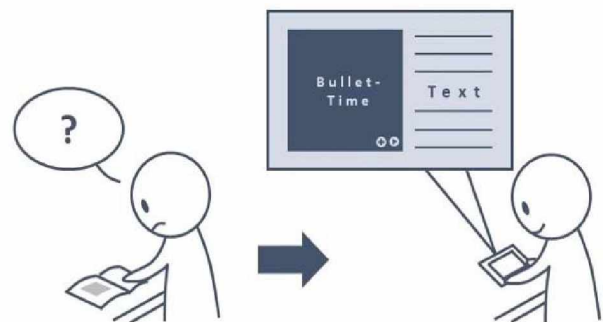
This paper proposes an interactive e-book with text and multi-perspective images to improve the readability of books. With the development of visual media technology, the expression and reading methods of book are diversifying. While text and images are the main methods of expression in conventional paper media, electronic books can be used for expression and viewing, which are difficult to achieve in paper media. In particular, electronic books, such as catalogs, illustrated books, and photo books, which are mainly used for the transmission of visual information, will be able to express richer contents by visual media technology.

If the shape of the observed object is planar, the expression of visual information is sufficient even from a single viewpoint (i.e., photo-media). On the other hand, when we observing a 3D object such as a statue or sculpture in detail, it may be difficult to understand the entirety by observing it from a single viewpoint. Therefore, it is often observed from multiple angles by moving the viewpoint position. To render the view from various perspectives using 3D-CG models has been applied for such observation. However, when we deal with a subject that requires more complex and photorealistic expression, the explanation by CG model is not sufficient. Bullet-time video has been used for generating 3D view with high realism and realize 3D perception based on motion parallax by arranging multiple cameras around a subject and observing multi-view images while switching between them sequentially [1][2].

Multimedia books have the effect of enabling the reader to understand the content of the book more deeply by allowing the viewer to manipulate the video and 3D model dynamically. Aristo [3] has succeeded to improve the readability of interactive e-books by using a camera mounted on an HMD to detect the reader's behavior, such as turning the pages of an e-book, and manipulating the information presented according to the detected results.

Our research aims to realize a deeper understanding of the content of the digital media by combining the Bullet-time video and the text information, as shown in Fig. 1. We aim to improve the browsability of the digital book contents by

utilizing bullet-time images to realize interactive operation where images and explanations interact with each other while expressing 3D visual information of the subject.



**Fig. 1: Interactive e-book linking of text and multi viewpoint video.**



**Fig. 2: e-Book Linking Text and Multi-View Video.**

## II. E-BOOK LINKING TEXT AND MULTI-VIEW VIDEO

This section explains the processing flow of the proposed method. We take multi-view images, and estimate the parameters of each camera by applying camera calibration (e.g., our system employees Structure from Motion (SfM) [4]). Using the estimated camera parameters, a projection transformation is calculated so that the optical axes of the multi-view cameras intersect at a single point in the 3D space, and a bullet-time video is generated by switching the multi-view images according to the camera arrangement. It is expected to improve the viewing effect by matching the point where the optical axis of the multi-view intersects with the point that attracts attention during viewing (gazing point). The creator of an e-book sets the gazing point at the 3D position where let the reader pay attention. As shown in Fig.2, we realize an interface of the interactive e-book by guiding the user's eyes through the reading function that is linked with the multi-view image and text. For example, by manipulating the description, such as highlighting and pointing, the viewpoint



of the multi-view image is switched, and the video information mentioned in the description is guided and presented to the user. In addition, the explanatory text is highlighted depend on the observing view point.

### III. DEVELOPMENT OF OUR INTERACTIVE E-BOOK

#### A. Shooting Multi-View Images and Bullet-Time Video Generation

Generally, multiple cameras are arranged evenly around a subject for capturing a bullet-time video. And the viewpoints are switched in order to display the image. If the subject is relatively small, we can place it on a turntable for shooting. However, it is difficult to take pictures of relatively large subjects in such way, so as shown in Fig.3, we shoot a subject while moving around it. In this case, it is necessary to compensate for differences in appearance caused by variations in the positional relationship between the camera and the subject. We estimate the camera parameters using SfM [4], and correct the appearance based on the parameters as shown in Fig.4. Considering the accuracy of camera calibration by SfM, it is better to set the camera interval (the convergence angle) to less than 10 degrees.

We developed a multi-view image browsing API based on bullet-time video "AnDonuts [5]" in collaboration with Denq Vision Inc. The generated bullet-time videos are stored on a cloud server instead of inside a user's terminal, and can be distributed as needed to enable embedding in an Internet-based e-book interface which we aim to develop.

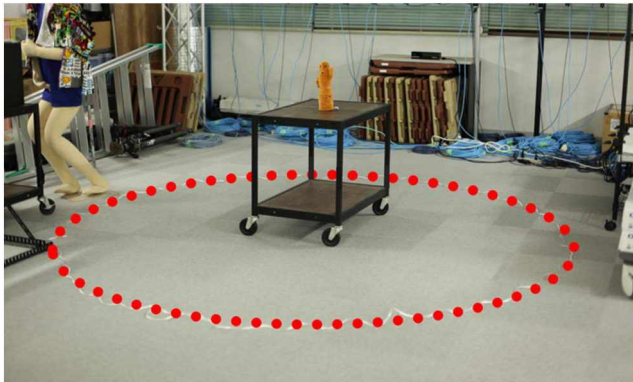


Fig. 3: Shooting environment (each camera is set on a red dot).

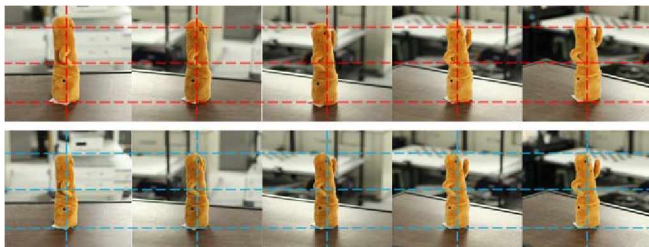


Fig. 4: An example of generated bullet-time video (Top: Input images, Bottom: Result images).

#### B. User Interface Implementation

Our browsing interface employs a framework "Django" to manage the web server and client. As a front end, the user interface is designed using HTML, CSS, and JavaScript, and the operation functions are implemented. When a web page

for viewing an e-book is loaded, the multi-viewpoint video presentation function is called via the API and the specified bullet-time video is presented. As shown in Fig.5, the user interface is divided into two areas: the bullet-time video is shown in the left area and the text about the work is shown in the right area. We employ the following operations to browse the bullet-time video based on our knowledge derived from the previous investigations [4].

- Viewpoint switching: Stroke the screen with mouse or touching screen.
- Zoom in/out: Rotate mouse wheel or pinch in/out with two fingers.
- Automatic viewpoint switching: Press Play button.

#### C. Interaction between text and multi-view images

##### 1) Controlling multi-view display by text manipulation

In books whose main purpose is to convey visual information, such as catalogues and illustrated books, the visual features of the subjects in the images are often explained in supplementary sentences. In our system, as illustrated Fig.5, when a user clicks on a particular keyword of interest while reading a description (e.g., Keyword A), the multi-view is switched to a viewpoint (VP-A) that helps the user understand the keyword. For example, in the case of "Haniwa" shown in Fig.4, if you click on the keyword "Through Hole", the viewpoint will move from the current one to where it is easy to observe the hole.

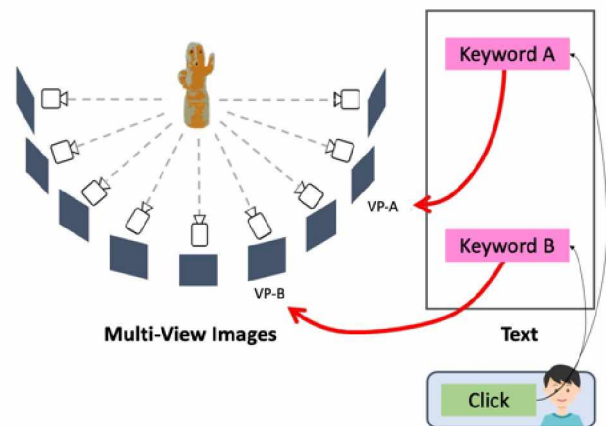


Fig. 5: Multi-view display through text manipulation.

##### 2) Controlling text display by multi-view manipulation

Switching perspectives changes the visual information the reader receives. Visualize the newly surfaced information by highlighting the sentence corresponding to that information. For example, as shown in Fig.6, when the viewpoint is switched to Haniwa viewpoint C (VP-C), the related description of the feature in the VP-C (e.g., "open hole") is highlighted in the text.

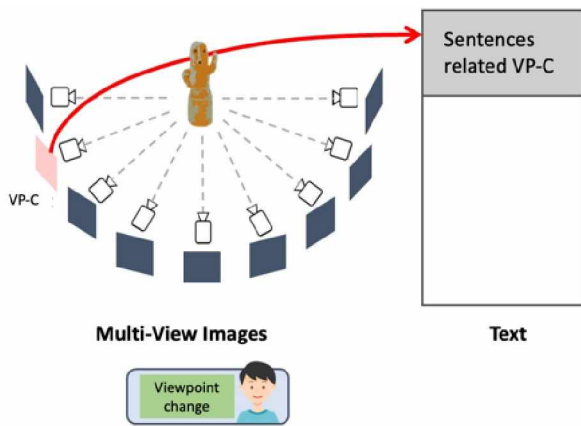


Fig. 6: Text display by viewpoint manipulation.

#### IV. PRACTICAL EXPERIMENTS

##### A. Shooting Experiment

In this section, we introduce a case study of the implementation for an exhibition at the Tokyo National Museum. In the archaeological exhibition room on the first floor of the Heiseikan, clay figures (dogu) from the Jomon period, copper bells (dotaku) from the Yayoi period, haniwa (clay figurines) from the Kofun period, and other artifacts are displayed. We captured multi-view images of eleven of the exhibits and generate each bullet-time video. The equipment used for the shooting was a Canon EOS 5D Mark II camera with an EF24mm f/2.8 IS USM lens and a SAMYANG AF14mm f/2.8 EF lens (resolution: 5616 pixels x 3744 pixels). Depending on the size of the subject and the size of the exhibition hall, a lens with an appropriate focal length was selected. The shutter speed, aperture value, color balance and other parameters were fixed so that the brightness and color temperature would not be affected by the position and orientation of the camera and light source. In order to reduce the specular reflection of the exhibit surface or the protective glass case, a polarization filter was installed on the lens to remove the reflection area to reduce the effect of reflection.

##### B. Museum Exhibit Implementations

An illustrated book or catalogue that presents artifacts with photographs of the exhibit also includes information about the exhibit by text, such as the location, region, and appearance characteristics of the artifacts that appear in the captions of the museum exhibit. We introduce examples of two patterns of implementation using the text information given by the book and the exhibition hall descriptions.

In the implementation of "Haniwa", 31 multi-viewpoint images were taken for about 240 degrees, and the description of the exhibition was given as Text. As shown in Fig. 7, key words such as "costume" and "Shimada topknot" are set as pink links in the text and multiple viewpoints.



Fig. 7: "Haniwa": Feature is marked with white point.

In the implementation example of the "Shakokidogu", multiple viewpoint images that can be observed from all surroundings are taken, and the text is used as an introduction to the exhibit. Here, a function to emphasize the explanatory text by switching the viewpoint of the video observation is implemented. When there is text information related to the displayed viewpoints, the text related to the features in the viewpoints is highlighted by pressing the upper left tag, as shown in Fig. 8.



Fig. 8: "Shakokidogu": Relevant article is highlighted with pink background.

##### C. Evaluation Experiments

###### 1) Experimental environment

To verify the effectiveness of the proposed method, we conducted an evaluation experiment on our e-books. The participants were 24 undergraduate and graduate university students in their twenties, who were divided into two groups according to the two types of the implementation mentioned in the previous subsection. The interface was implemented on a Microsoft Surface Pro 4 (CPU: Intel(R) Core(TM) m3-6Y30, memory: 4GB). In the experimental procedure, the participants start reading the book after being confirmed their understanding of the e-book operation, and answer a questionnaire at the end of the reading. Because of individual differences in browsing efficiency, the time required for browsing was left to the decision of each participant.

###### 2) Experimental procedure

The questionnaire consists of two parts. In the first part, two simple tests Q1 and Q2 were conducted to objectively



examine whether our proposed e-book can improve the level of understanding.

In the second part, we ask questions Q3 to Q7 about the subjective evaluation of the experimental participants' operations. The questions are listed below.

- Q3: Is it possible to get 3D shape information of the subject through observing the multi-view images?
- Q4: Are you satisfied with the smoothness of viewpoint switching?
- Q5: Is it possible to observe the subject with fully understanding the function that allows text and multiple viewpoints to work together?
- Q6: Does the function of linking text and multi-view improve the efficiency of reading a book?
- Q7: Does the function of linking text and multi-view facilitate your understanding of the content?

In Q3, we confirm that observations from multiple perspectives provide some 3D information. Q4 aims to check whether the number of inputs for multiple viewpoints is sufficient. Q5 aims to confirm the operability of the function. Q6 aims to confirm that the viewing efficiency has been improved. And in Q7, we confirm that the understanding of the text has improved. The participants answer the questions by rating on a 5-point Likert scale. ("5: yes", "4: somewhat yes", "3: neither", "2: somewhat no", "1: no").

### 3) Experimental results

#### a) Objective evaluations

The results of the questionnaire in the first part (Q1, Q2) are explained in this section. About "effect on controlling multi-view display by text manipulation", All participants who read our e-book and a general book answered Q1 and Q2 correctly. About "effect on controlling text display by multi-view manipulation", all the participants answered correctly in Q1, and 11/12 participants who read our e-book and 10/12 participants read a general book answered Q2 correctly.

As the result of this experiments, it is difficult to confirm the improvement of understanding by the introduction of our e-books, because the majority of the participants answered correctly. This may be due to the fact that the sentences used in the experiment were short and the number of questions was small due to the limited number of characters displayed on the screen. It is necessary to increase the difficulty of the questionnaire by using longer sentences and giving more questions.

#### b) Subjective evaluations

The results of the questionnaire in the second part (Q3 to Q7) are explained below. Fig. 9 shows the results for an e-book equipped with the "Effects of controlling multi-view display by text manipulation" function. All participants evaluated positively the ability to express 3D information in multi-view images (Q3), the feeling of viewpoint switching by the number of inputs (Q4), and the feeling of operation (Q5). The opinions about the improvement of reading efficiency (Q6) differed according to the participants' reading habits. While some participants said that "looking at the text and the images interchangeably sometimes interrupted the reading of the text," others said that "the efficiency of reading the text was reduced in terms of speed, but increased in terms of reading with comprehension". With regard to the promotion of understanding of the content (Q7), one of the participants responded that "many people are more interested

in the photographs because the text and the multi-view images work together, and many of them want to look at the pictures, even if they are not explained".



**Fig. 9: Experiment result of "Controlling multi-view display by text manipulation". ("sky blue: yes", "yellow: somewhat yes", "gray: neither", "orange: somewhat no", "navy blue: no").**

Fig. 10 shows the results for an e-book with "Effects of controlling text display by multi-view manipulation" function. All participants gave positive evaluations of the ability to represent 3D information in multi-view images (Q3), the operability of this function (Q5), and the promotion of understanding of the content (Q7). As for the feeling of viewpoint switching by the number of input multiple-viewpoint images (Q4), we were not able to switch viewpoints smoothly due to the omission of recognition of the tablet's touch operation, and as a result, the evaluation score is not so high. We asked the same participants to read the same e-book again with using a PC (mouse operation), and we confirmed that the above-mentioned problems were solved and the evaluation was improved. With regard to the improvement of viewing efficiency (Q6), one said that "I was able to learn information efficiently because I could observe from multiple viewpoints" and the other said that "The highlighted text linked to the images is long. It seems to improve the learning efficiency if the text is highlighted word by word".



**Fig. 10: Experiment result of "Controlling text display by multi-view manipulation". ("sky blue: yes", "yellow: somewhat yes", "gray: neither", "orange: somewhat no", "navy blue: no").**

### 1) Discussions

The results of the subjective evaluations show that our proposed e-book enables 3D observing the subject by displaying multi-view images and promotes the feeling of operation and understanding of the content. From the experimental results on whether the number of inputs is sufficient for the multiple-viewpoint video, it was found that individual parameter adjustments are necessary for each terminal (browsing condition). In the evaluation of the efficiency of browsing, the results differed according to the participants' browsing habits, suggesting that it is necessary to improve the functions of the system according to the participants' age and situation of use.

## V. CONCLUSIONS

This paper have proposed an interactive e-book with multi-view images and text to improve the readability of books, which are mainly intended to convey visual information. Specifically, our interactive e-book has a viewing function in which the viewpoint of the multi-view images and the explanatory text interact with each other by implementing (1) a function to change the viewpoint of the multi-view images by manipulating the explanatory text such as highlighting and pointing, and (2) a function to highlight the explanatory text that enhances the understanding of the view of the multi-view images according to the viewing viewpoint.

Through demonstration experiments of the proposed method for museum exhibits, we confirmed that generating

of bullet-time video" and our e-books can be realized at a practical level. As a result of the evaluation experiments, it was confirmed that the proposed e-book can promote the ability to express three-dimensional information and the comprehension of the content. If the browsing function is further improved based on the user's browsing habits, it is possible to be applied to various situations.

This work was partly supported by "Research Support Program to Tackle COVID-19 Related Emergency Problems, University of Tsukuba" and "Grants-in-Aid for Scientific Research (17H01772)".

## REFERENCES

- [1] Nao Akechi, Itaru Kitahara, Ryuuki Sakamoto, Yuichi Ohta, "Multi-Resolution Bullet-Time Effect", SIGGRAPH Asia 2014 Posters Article No.30, 2014.
- [2] Nobuyuki Kitamura, Hidehiko Shishido, Takuya Enomoto, Yoshinari Kameda, Jun-Ichi Yamamoto, Itaru Kitahara, "Development of Multi-View Video Browsing Interface Specialized for Developmental Child Training", Proceedings of Asia Pacific Workshop on Mixed and Augmented Reality (APMAR2019), pp.1-8, 2019.
- [3] Zhongyang Zheng, Bo Wang, Yakun Wang, Shuang Yang, Zhongqian Dong, Tianyang Yi, Cyrus Choi, Emily J. Chang, Edward Y. Chang, "Aristo: An Augmented Reality Platform for Immersion and Interactivity", ACM Multimedia Conference, 690-698, 2017.
- [4] C. Wu, "Towards Linear Time Incremental Structure from Motion", Proceedings of the 2013 International Conference on 3D Vision, pp.127-134, 2013.
- [5] "AnDonuts", <https://denqvision.com>